



Publication

CA24N
DG
-76527
v.2

Secondary-Postsecondary Interface Project II:

Nature of Students



VOLUME 2

By

ROSS TRAUB, RICHARD WOLFE, CARLA WOLFE,

PETER EVANS, H.H.RUSSELL

Summary by

KATHERINE WAYNE

This research project was funded under contract by the Ministry of Education,
and the Ministry of Colleges and Universities, Ontario.

SECONDARY-POSTSECONDARY INTERFACE
PROJECT II:
NATURE OF STUDENTS

by

Ross Traub, Richard Wolfe, Carla Wolfe,
Peter Evans and H.H.Russell

with the assistance of

Marc Barnes, Monique Bélanger, Michael Canale,
Carol Dimond, John Hattie, Marion Marsh
and Raymond Mougeon

Summary by Katherine Wayne

VOLUME TWO

This research was funded under contract by the Ministry of
Education and the Ministry of Colleges and Universities,
Ontario

TABLE OF CONTENTS

APPENDIX A1:	TECHNICAL REPORT ON THE TESTS OF ENGLISH FOR ANGLOPHONES.....	1
1.	TEST SELECTION AND TEST CONTENT....	1
2.	APPRAISAL OF THE TESTS.....	5
3.	TECHNICAL ISSUES.....	37
4.	STYLE (DICTION).....	66
5.	FLAVOUR.....	67
	REFERENCES.....	94
APPENDIX A1A:	SCORING OF ESSAYS.....	95
APPENDIX A1B:	THE ESSAY ERROR COUNT....	98
APPENDIX A1C:	INSTRUCTION SHEET FOR ESSAY APPRAISALS.....	106
	TABLES A1.1 TO A1.9.....	108
APPENDIX A2:	TECHNICAL REPORT ON THE TESTS OF FRENCH FOR FRANCOPHONES.....	119
1.	DESCRIPTION OF TEST CONTENT.....	119
2.	TEST APPRAISAL.....	121
3.	TECHNICAL ISSUES.....	123
	REFERENCES.....	132
	TABLES A2.1 TO A2.10.....	133
APPENDIX A3:	TECHNICAL REPORT ON THE TESTS OF FRENCH AS A SECOND LANGUAGE.....	159
1.	DESCRIPTION OF TEST CONTENT.....	159
2.	TEST APPRAISAL.....	162
3.	TECHNICAL ISSUES.....	177

REFERENCES.....	194
TABLES A3.1 TO A3.30.....	195
APPENDIX A4: TECHNICAL REPORT ON THE <u>TEST DE CONNAISSANCE DE LA LANGUE (ANGLAIS)</u>	237
1. DESCRIPTION OF TEST CONTENT.....	237
2. TEST APPRAISAL.....	238
3. TECHNICAL ISSUES.....	245
REFERENCES.....	251
TABLES A4.1 TO A4.6.....	252
APPENDIX A5: TECHNICAL REPORT ON THE <u>PHYSICS ACHIEVEMENT TEST</u>	269
1. DESCRIPTION OF TEST CONTENT.....	269
2. TEST APPRAISAL.....	273
3. TECHNICAL ISSUES.....	280
REFERENCES.....	284
TABLES A5.1 TO A5.5.....	285
APPENDIX A6: TECHNICAL REPORT: <u>MATHEMATICS ACHIEVEMENT TEST</u>	301
1. DIFFICULTY.....	303
2. BISERIAL CORRELATION.....	304
3. RELIABILITY.....	305
4. SPEEDEDNESS.....	305
5. CONTENT VALIDITY.....	306
TABLES A6.1 TO A6.5.....	311

APPENDIX A7:	TECHNICAL REPORT: <u>TEST OF ARITHMETIC AND BASIC ALGEBRA</u>	321
	1. DIFFICULTY.....	322
	2. BISERIAL CORRELATION.....	322
	3. RELIABILITY.....	322
	4. SPEEDEDNESS.....	322
	5. TEST SCORES.....	323
	6. CONTENT VALIDITY.....	323
	TABLES A7.1 TO A7.6.....	328
APPENDIX A8:	TECHNICAL REPORT ON THE <u>TEST DE RENDEMENT EN MATHEMATIQUES</u>	341
	1. TEST CONTENT.....	341
	2. TECHNICAL ISSUES.....	341
	3. TEST APPRAISAL.....	343
	TABLES A8.1 TO A8.5.....	347
APPENDIX A9:	TECHNICAL REPORT ON THE <u>TEST D'ARITH- METIQUE ET D'ALGEBRE DE BASE</u>	357
	1. TEST CONTENT.....	357
	2. TECHNICAL ISSUES ABOUT THE TEST....	357
	3. TEST APPRAISAL.....	359
	TABLES A9.1 TO A9.5.....	363
APPENDICE B:	ANALYSE DETAILLEE DES TRAVAUX ECRITS D'UN ECHANTILLON D'ELEVES FRANCO-ONTARIENS DE 12EME ET 13EME ANNEES.....	373
	1. INTRODUCTION.....	373
	2. PRESENTATION DE L'ECHANTILLON.....	375
	3. METHODOLOGIE.....	376
	4. RESULTATS.....	378

5.	CONCLUSION.....	391
	TABLEAUX B.1 TO B.9.....	393
	GRAPHIQUE B.1.....	402
APPENDIX C:	BASIS FOR CLASSIFYING UNIVERSITY COURSES.....	403
APPENDIX D1:	EQUATING SCORES ON BOTH FORMS OF THE TEST OF READING COMPREHENSION AND <u>LANGUAGE ACHIEVEMENT (ENGLISH) AND</u> <u>THE TEST DE COMPREHENSION EN LECTURE</u> <u>ET DE CONNAISSANCE DE LA LANGUE</u> <u>(FRANCAIS).....</u>	405
	1. PROCEDURE.....	405
	2. APPLICATIONS.....	406
	TABLES D1.1 TO D1.2.....	408
APPENDIX D2:	PROCEDURE FOR SCORING ESSAYS.....	410
	TABLE D2.1.....	415
APPENDIX E:	STUDENT QUESTIONNAIRE.....	417
APPENDIX F:	TABLES OF STATISTICS REFERRED TO IN THE SUBSECTION ON MARKING STANDARDS (CHAPTER 3, 1.3) OF THE SSGD/SSHGD SURVEY OF ANGLOPHONE STUDENTS.....	435
	TABLES F.1 TO F.7.....	436
APPENDIX G:	TABLES OF STATISTICS REFERRED TO IN THE SUBSECTION ON MARKING STANDARDS (CHAPTER 3, 2.2) OF THE SSGD/SSHGD SURVEY OF FRANCOPHONE STUDENTS.....	445
	TABLES G.1 TO G.7.....	446

APPENDIX H:	TABLES OF STATISTICS REFERRED TO IN THE REPORT OF RESULTS FROM THE UNIVERSITY SURVEY (CHAPTER 3, 3.1).....	461
	TABLES H.1 TO H.8.....	462
APPENDIX I:	TABLES OF STATISTICS FOR THE STUDY OF FACTORS RELATED TO SCHOOL ACHIEVEMENT (CHAPTER 3, 1.4).....	479
	TABLES I.1 TO I.8.....	480



Digitized by the Internet Archive
in 2024 with funding from
University of Toronto

<https://archive.org/details/31761118918184>

APPENDIX A1

TECHNICAL REPORT ON THE TESTS OF ENGLISH FOR ANGLOPHONES

This report is divided into three main sections. The first section contains a description of the contents of the English tests for Anglophones together with mention of some of the considerations that shaped the design of the tests. The second section consists of an explanation of the procedure by which the tests were appraised, and a presentation and discussion of the results of that appraisal. The third section deals with some technical matters related to the tests.

1. TEST SELECTION AND TEST CONTENT

The Advisory Committee charged with the task of assisting in selecting or designing appropriate instruments for evaluating the English language competency of Anglophone students took as their aim the testing of reading, language, and writing. The test package ultimately prepared focussed on those three major components of literacy at a level of difficulty appropriate for students approaching the Interface. None of the tests addressed themselves to specific Grade Twelve or Grade Thirteen course objectives.

1.1 Testing Reading

Available multiple-choice reading tests were searched for passages which addressed reading comprehension in some depth, instead of settling for something rather closer to recognition of surface

detail. Questions which demanded a variety of skills (reading for main idea or purpose, and reading to see implications and make inferences, in addition to reading for literal meaning) were seen as ideal. Other goals of the search were interest and variety in passage content, since the effect of those qualities on student readers' motivation--and therefore on test validity--was recognized.

Eventually four passages were drawn from the Service for Admission to Colleges and Universities bank of tests designed for university entrants. All four appeared in the Canadian Scholastic Aptitude Test (1973) as a subtest of scholastic aptitude, not as an English achievement test. Since they were designed for university-bound students in their last year at secondary school, it was anticipated that the passages would be rather too difficult for most general level Grade Twelve students, but, for purposes of comparing Grade Twelve and Grade Thirteen performances, one set of passages had to be selected for both groups of students.

1.2 Testing Language Achievement

Two subtests were selected for testing the language component of literacy: (i) the "construction shift" set employed in the Canadian Test of English Language Form 760 (1974), and (ii) the "sentence correction" set, with minor emendations, employed in the Canadian English Language Achievement Test Form 322 (1970), using the operational items only.

In the "construction shift" type of exercise, the student is presented with a sentence which is grammatically correct. He is directed to make a specific change in the sentence, and then, working within the limitations dictated by that specific change, he is to rephrase the entire sentence, retaining the original meaning as far as possible. His revised sentence must contain one of five optional words or phrases. These are labelled (A) to (E), and the student marks the one incorporated in his revision of the original sentence.

In the "sentence correction" type of exercise, the student is given a sentence in which a section is underlined. Options (A) to (E) show five ways of phrasing the underlined section, including--as (A)--the original way. The student is to select the clearest and most correct of the five phrasings.

These two types of language exercise were selected by the English Advisory Committee for inclusion in the test package because of their high correlation with total essay scores, as reported by Godshalk, Swineford, and Coffman (1966). At the same time, the Committee recognized that performance on any form of multiple-choice language test--even a form like the "construction shift", demanding sentence manipulation--was a very dubious approximation of the student's performance with language. Hence the insistence upon a test of student writing.

1.3 Testing Writing

The Advisory Committee realized that, to ensure reasonable reliability and validity, the ideal writing test (Britton, Martin and Rosen, 1966) would consist of essays in two modes, with no topic choice, to be written on different days. The writing test eventually included in the English package "involved some compromises with this ideal.

- (a) Test administration constraints allowed only one-half day for testing writing. Thus it was possible to ask for only one essay.
- (b) In the interest of reliability the single essay was restricted to a single mode (expository: argument or viewpoint)--the mode judged by the Advisory Committee as receiving most emphasis in English Composition at the senior secondary level, and most important for postsecondary success regardless of the student's area of specialization.

- (c) Several topics were offered. Though provision of a wide choice was recognized by the Advisory Committee as inevitably reducing scoring reliability, it was intended, on balance, to increase the writing test's validity; presumably a student would be more motivated to make a serious effort at arguing his viewpoint if the subject under discussion permitted him to choose a topic of some interest to him.

1.4 The Test of Reading Comprehension and Language Achievement

There were two parallel 40-minute forms of this test, designated as Form 1 and Form 2. Each form contained two reading comprehension passages with five related items per passage. An analysis of the items (combining Forms 1 and 2) ran as follows:

- (a) items dealing with literal meaning - 10;
- (b) items dealing with main idea (implied) - 2;
- (c) items dealing with inference, implication - 7;
- (d) miscellaneous items - 1;

Each form also contained eleven "construction shift" items followed by fifteen "sentence correction" items.

Test administration conditions required some Grade Thirteen physics students to write the physics test in the second half of the morning testing period. Also in that time period, a sub-sample of students wrote the essay. Students writing neither the essay nor the physics test took both forms of The Test of Reading Comprehension and Language Achievement; those writing physics or the essay were randomly assigned to either Form 1 or Form 2.

1.5 The Writing Test

The sub-sample of Grade Twelve and Thirteen Anglophone students writing English tests were given 1-1/4 hours to write a 250-350 word essay on one of eight topics. They were given the following instructions:

Develop your position with reference to the topic chosen, supporting it with carefully chosen illustrations and carrying it forward to a logical conclusion. Feel quite free to agree, disagree, or take an intermediate position with respect to the point of view stated or implied by the topic.

2. APPRAISAL OF THE TESTS

Both forms of the Test of Reading Comprehension and Language Achievement (English) and the Writing Test were sent to teachers of Grade Twelve and Grade Thirteen English courses in the 53 Anglophone schools in the study and to a selected number of the instructors of first year English courses in 15 Colleges of Applied Arts and Technology and in 11 universities. In addition, each of these individuals was sent an appraisal inventory for the tests. Secondary school teachers were asked to appraise the tests with reference to students taking the following English courses: Grade Thirteen, Grade Twelve Advanced, Grade Twelve General and Grade Twelve Basic. Postsecondary teachers were asked to respond to the inventory with reference to first year regular and/or remedial English programs.

The secondary school edition of the inventory consisted of forty questions and the postsecondary edition of thirty-two. All thirty-two of the questions on the postsecondary edition appeared also in the secondary edition with certain significant differences in phrasing. For instance, where secondary instructors were asked to consider student performance upon completion of their courses,

postsecondary instructors were asked to consider the performance of students entering their first year courses. For ease of comparison, the responses of instructors at both levels will be reported here in parallel.

The additional questions included in the secondary school inventory referred in detail to specific passages in the reading test. Only overall impressions were solicited from postsecondary teachers.

All questions called for response by number for coding and computer print-out; frequent opportunities to supplement coded information with written responses were provided. Teachers could respond in the code columns for the one, two, three or four courses they taught, and then add comments appropriate to one or more of the courses. The comments were read and summarized by Project II staff following the keypunching of the coded responses.

2.1 Nature of the English Inventory

The following limitations and purposes of the appraisal inventory should be clarified at the outset:

- (a) Because testing was limited only to reading, language, and writing competency (the last restricted further to only one mode), the test inventory was similarly restricted. Many important areas of program had to be left aside: for example, spoken English and literature.
- (b) The question of how best to test competences was a major concern of the inventory, every bit as important as the question of appropriateness of objectives tested or of difficulty level. It was hoped that the appraisals might throw some helpful light on issues much debated among teachers, such as the utility of multiple-choice English tests or reliability in evaluating writing samples.

- (c) The English appraisal inventory did not line up individual items from the multiple-choice tests against individual specific course objectives, inquiring, "Did you teach this?" in the way that mathematics or physics appraisal inventories could and did. Most items testing language--and reading too, to a considerable extent--did not test for a single skill. Instead, the sum of items was a sort of "blitz" of these areas of literacy, and the inventory dealt with them as such.

The following example of a "sentence correction" item illustrates the combination of language skills required by such items:

The climax is when the hero stabs the villain.

- (A) is when the hero stabs the villain
- (B) is where the hero stabs the villain
- (C) occurs when the hero stabs the villain
- (D) occurs when the hero stabbed the villain
- (E) is the hero has stabbed the villain

The selection of the correct answer, "C", depends not only on the awareness that "when" and "where" are inappropriate conjunctions for beginning the noun clause (eliminating choices A and B) but also on an awareness of tense sequence: "occurs...stabs" rather than "occurs...stabbed". While the combination of otherwise unrelated grammatical concerns does not render the test item inappropriate, it does make link-up between test item and a specific language objective very difficult. The student who selects (D) or (E) is partly right in that we may assume he has identified and avoided one type of error, but his being wholly right depends upon awareness of a second type of

error. It should be noted as well that the practice item used here for illustration is considerably simpler to unravel than many of the more complex items in the actual test.

2.2 Responses

Altogether there were 136 respondents for Grade Thirteen, 115 for Grade Twelve advanced, and 93 for Grade Twelve general. Only 11 teachers responded for Grade Twelve basic level; therefore, in instances where a report of these 11 responses, given as a percentage, would lend the appearance of far more precision and consensus than the data warrant, "basic" level responses have been deleted.

Teachers were requested to respond to the inventory for each grade and level taught. Consequently many responded for two or three levels, and provided a mixture of written responses, some generally applicable and some specific to a level.

A total of 34 CAAT and 25 university teachers responded to the appraisal inventory. They were asked to respond for "regular" first year or "remedial" or both, depending upon the courses they taught. However, as only two university respondents indicated any involvement with remedial programs, their responses were combined with other university first year responses.

There appeared to be some confusion on the part of postsecondary appraisers concerning response columns "regular" and "remedial"; consequently there was some crossing of data which was difficult to sort out. Since university responses were recorded as one total, the confusion is not an issue there, but it does remain a problem in interpreting CAAT responses with desirable precision. Total responses for CAAT regular English programs were 28 and, for remedial, 16.

2.3 Interpreting Recorded Responses

In view of the numbers responding for secondary schools and in an attempt to facilitate a reasonably straightforward comparison between viewpoints of secondary and postsecondary teachers, responses have usually been recorded as percentages, except where means (as in assessment of difficulty by use of a scale) seemed more appropriate. Unless otherwise noted, then, answers in the following tables are expressed as percentages.

Because there were, relatively, so few postsecondary responses, and because the postsecondary instructors who were invited to respond were chosen on arbitrary grounds and as a group cannot be said to represent a probability sample of postsecondary instructors across the province, one must be careful not to interpret percentage scores with a precision and generality they do not deserve. It would be inappropriate to consider percentage here as any firm consensus of postsecondary teachers of English or to judge differences as between 80 per cent and 85 per cent as being particularly significant.

The order of questions as they appeared in the inventory has been re-arranged somewhat in this report. Response columns have been designated as follows:

- 13 - Grade 13;
- 12A - Grade 12 Advanced;
- 12G - Grade 12 General;
- 12B - Grade 12 Basic;
- CAAT - First year CAAT ("Regular" and "Remedial" are split when appropriate);
- UNIV- First year of university.

2.4 Reading Comprehension

It is apparent from the responses to the following eight items that both postsecondary and secondary school teachers across the

grades and levels involved saw the testing of reading comprehension as a very important component in the evaluation of language achievement. Many saw competence in deriving literal meaning, deriving main idea, and deriving inference or implication as important on entry to the grade level and as important objectives in the program.

Do you consider that a test of reading comprehe- hension assesses an important com- ponent of language achievement?	13	12A	12G	12B	CAAT	UNIV
Yes	95	96	92	93	94	100

How many of the students <u>entering</u> courses at this level should have the ability to read a passage for <u>literal</u> <u>understanding</u> ?	13	12A	12G	12B	CAAT	UNIV
100%	73	61	47	15	72	83
76%-99%	19	18	19	38	19	17
51%-75%	5	15	18	23	3	0
26%-50%	2	3	7	8	6	0
1%-25%	0	3	4	8	0	0
None	0	0	0	0	0	0

How many students
on entry should
be able to identi-
fy the main idea
or purpose of
a passage?

	13	12A	12G	12B	CAAT	UNIV
100%	65	51	38	23	63	83
76%-99%	27	33	28	31	22	17
51%-75%	7	12	22	31	16	0
26%-50%	1	3	4	0	0	0
1%-25%	0	1	4	15	0	0
None	0	0	3	0	0	0

How many students
on entry should
be able to draw
inferences and
see implications?

	13	12A	12G	12B	CAAT	UNIV
100%	27	15	8	8	31	42
76%-99%	37	20	20	17	19	38
51%-75%	29	38	36	25	31	20
26%-50%	3	23	16	33	9	0
1%-25%	4	3	16	17	9	0
None	0	1	3	0	0	0

What emphasis in
this course is
given to reading
for literal
understanding?

	13	12A	12G	12B	CAAT	CAAT	UNIV
						REG.	REM.
heavy	37	34	49	23	44	50	36
moderately heavy	24	34	32	54	31	25	20
light	26	22	16	23	13	0	24
remedial only	13	9	2	0	13	25	12
none	0	3	1	0	0	0	8

What emphasis in
this course is
given to reading
to identify the
main idea or
purpose?

	13	12A	12G	12B	CAAT	CAAT	UNIV
						REG.	REM.
heavy	57	53	62	76	53	67	52
moderately heavy	33	36	32	33	34	17	40
light	7	10	4	0	6	8	0
remedial only	2	1	2	0	6	8	4
none	1	0	0	0	0	0	4

What emphasis in
this course is
given to drawing
inferences and
seeing impli-
cations?

13	12A	12G	12B	CAAT	CAAT	UNIV
					REG.	REM.

heavy	64	41	16	8	25	8	64
moderately heavy	30	51	41	42	44	50	28
light	5	7	33	33	28	33	4
remedial only	0	1	8	17	3	8	4
none	1	0	1	0	0	0	0

In response to a further question, "Are there important reading skills that were not tested but that should have been?" a high proportion responded affirmatively: University - 50 per cent; CAAT - 29 per cent; 13 and 12A - 40 per cent; 12G - 36 per cent. The range of suggestions was quite wide. The four appearing with greatest frequency were vocabulary; inference or implication (perhaps implying that the test items did not go far enough in measuring those skills); critical evaluation of argument (for structure, use of evidence, logic, bias, etc.); and appreciation (presumably of literary qualities). Only about ten of those responding suggested testing for rate.

It is reasonable to conclude that, in the opinion of both secondary and postsecondary teachers, the test of reading comprehension attacked quite appropriate objectives, though it would have been desirable to expand and deepen them somewhat, focussing on critical awareness of specific aspects of language and style.

Regarding the suitability of reading passages, secondary teachers were asked both general questions and questions about

each passage; postsecondary teachers were asked the general questions only.

From the standpoint both of expectations concerning type of material students should be able to read and of appropriateness of level of difficulty, there was a high level of agreement between Grade Thirteen and university teachers and between CAAT and Grade Twelve advanced teachers. Teachers of Grade Twelve general students considered the passages too difficult and teachers of Grade Twelve basic level considered the passages far beyond the students' range.

Are the four pas-
sages representa-
tive of the mater-
ial you would
expect students
at this level to
be able to read
with comprehen-
sion?

	13	12A	12G	CAAT	UNIV
Yes	72	50	14	52	80

What is your over-
all assessment of
the difficulty
level of the four
passages given in
the two forms of
this test?

	13	12A	12G	12B	CAAT	UNIV
too easy	1	0	1	0	3	0
somewhat easy	5	3	1	0	3	0
about right	45	32	11	7	26	71
somewhat difficult	42	40	36	13	47	29
too difficult	8	25	51	80	21	0
	----	----	----	----	----	----
Mean:	3.50	3.86	4.35	4.73	3.80	3.29

Only the secondary school teachers were asked to respond concerning the specific passages. They were asked to make this assessment using the range from 1 (too easy) to 5 (too difficult). Their responses are reported here as means:

	13	12A	12G
Passage on Mass Marketing	3.14	3.59	4.10
Passage on Viviparity	3.39	3.72	4.17
Passage on Object-Perception	3.81	4.23	4.39
Passage on Mackenzie King	3.27	3.61	4.20

Is the passage appropriate except for difficulty?	13	12A	12G
(a) Mass Marketing			
Yes	82	81	73
(b) Viviparity			
Yes	58	56	61
(c) Object-Perception			
Yes	58	56	54
(d) Mackenzie King			
Yes	87	80	74

It is of interest that while most teachers found passages of sociological commentary ("Mass Marketing") and historical commentary ("Mackenzie King") appropriate, those on scientific topics were considered inappropriate by a substantial proportion of respondents. While a few made criticisms regarding style or remoteness from student interests, the strongest criticisms, almost exclusively restricted to "Viviparity" and "Object-Perception", were of the technical and abstract vocabulary which were considered as seriously disadvantaging the non-scientific student. It should be noted, however, that though criticism on this ground occurred in at least 30 per cent of the responses, score results indicate that students did slightly better on the "scientific" passages than on the other pair. Some teachers recommended passages from literature or passages of higher literary quality as alternatives to the "technical" ones.

The criticisms raise important questions. On tests of reading at the senior secondary level, is it appropriate or fair to include topics which are rather specialized? Should secondary education be training students to read materials in a wide range of topic areas? If so, should the responsibility fall principally on the teacher of English?

The postsecondary teachers, though they had not been invited to comment on specific passages, made a few general suggestions: that the range of passages be broader, and that controversial passages (such as "Mackenzie King", presumably) be avoided. A few noted that the present passages lacked interest.

Respondents were asked also to comment on the multiple-choice format for testing reading comprehension.

Is the multiple-choice format a reasonable method of assessing at least the three reading comprehension abilities mentioned?

	13	12A	12G	CAAT	UNIV
Yes	40	46	41	62	40
Yes (qualified)	53	49	50	33	48
No	7	5	9	4	12

Over 90 per cent of the respondents reacted favourably to the multiple-choice format for reading comprehension though approximately 50 per cent had serious qualifications. These concerned test conditions such as the time allowance (judged inadequate), testing pressures, or guessing as a factor; the wording of items; and certain inadequacies of the test. Its failure to measure more complex understandings or subtle distinctions was noted by some, and some expressed the need for a response from the student "in his own words". In fact, almost all of the criticisms pointed to a need for some written response on the part of the student.

Thus, one might conclude from the appraisal of the reading test that teachers of English at all levels would be reasonably satisfied with a test of reading comprehension combining multiple-choice and written responses and restricted to reading passages in general interest areas, not necessarily "English".

2.5 Language Achievement

Respondents at both levels were asked to consider the objectives tested by both the "construction shift" and "sentence correction" language exercises.

What is your evaluation
of the language
achievement items
in the tests from
the standpoint of
emphasis given to
usage, style,
grammar, struc-
ture, and idiom?

13	12A	12G	CAAT	UNIV
----	-----	-----	------	------

They provide a reas-
onable balance in
testing important
areas of language
achievement.

63	63	70	81	71
----	----	----	----	----

Though testing a number of import- ant areas, there is an imbalance in emphasis.	17	12	15	6	4
--	----	----	----	---	---

Important areas of language achievement are omitted or tested too lightly.	19	24	14	13	25
---	----	----	----	----	----

Those responding (2) or (3) referred less to balance and emphasis than to areas of language competence they felt should also have been tested. These complaints included the neglect of crude as opposed to subtle errors (e.g. sentence fragment) and the failure to test mechanics such as punctuation and spelling. But the infrequency of these criticisms leads to the conclusion that, overall, there was a strong general consensus that the test addressed itself to matters of language for which competence was expected for a large proportion of students by the end of the secondary program: 83 per cent responding for Grade Thirteen believed these competences should be achieved by 75 per cent or more of the students. The figure for Grade Twelve advanced was 72 per cent of the respondents, and for Grade Twelve general, 45 per cent. In comparison, 69 per cent of the CAAT teachers and 84 per cent of the university instructors expected these competences from 75 per cent or more of the students on entry to those institutions. (Please note: In all instances "competence" means competence in the areas of language tested and not necessarily near perfect performance on the test items.)

How many students
should have the
competences assessed
by the language
achievement items
upon entry to your
course? (Postsecondary:
"institution" replaces
"course ")

	13	12A	12G	CAAT	UNIV
100%	36	26	23	41	56
76%-99%	34	23	15	28	28
51%-75%	24	34	28	13	8
26%-50%	2	11	20	9	4
1%-25%	1	3	10	3	4
none	3	3	3	6	0

How many of the
students who
successfully
complete English
courses at this level
should have the
competences assessed
by the language
achievement items?

	13	12A	12G	CAAT REG.	CAAT REM.	UNIV
100%	44	37	17	28	38	70
76%-99%	39	35	28	31	31	13
51%-75%	12	18	31	24	6	9
26%-50%	1	4	13	10	12	0
1%-26%	1	0	9	0	6	4
none	3	5	3	7	6	4

Further, as the following table demonstrates, the test
focused on aspects of language considered important across the

Interface and therefore given considerable emphasis at every level.

How much emphasis
do you give to de-
veloping the com-
petences demanded
by the language
achievement items?

	13	12A	12G	CAAT REG.	CAAT REM.	UNIV
none--competences are unimportant	1	2	1	0	0	0
heavy	16	12	10	27	65	20
moderately heavy	43	50	45	50	12	12
light	22	13	15	13	12	56
remedial only	22	13	15	13	12	56
none--competences are too advanced	2	3	7	0	6	0

Respondents were asked to assess the two item types used in the language achievement test and the difficulty level at which they were set.

What is your eval-
uation of the "con-
struction shift"
items?

	13	12A	12G	CAAT	UNIV
Suitable but too easy	8	9	2	0	0
Suitable and right for difficulty	52	49	33	34	60

Suitable but too difficult	10	16	42	24	8
Unsuitable for reasons other than difficulty	29	27	23	32	32
What is your eval- uation of the "sen- tence correction" items?	13	12A	12G	CAAT	UNIV
Suitable but too easy	13	9	2	0	4
Suitable and right for difficulty	67	72	47	76	86
Suitable but too difficult	10	15	38	18	0
Unsuitable for reasons other than difficulty	9	4	12	6	9

Those responding with any qualification about the item types were encouraged to comment further:

Under "suitable", very few respondents had negative comments to make concerning the "sentence correction" type. Indeed, it received a number of gratuitous favourable comments.

"Construction shift", however, was not so favourably received. The format and instructions for this item type were considered by many from both secondary and postsecondary levels to be very confusing for students; a number of secondary school teachers claimed that the student could not determine what he was asked to do. The relatively high percentage of students failing to complete this section of the test adds weight to this criticism: the average completion per item for the construction shift by Grade Twelve students was 82 per cent, while for the sentence correction type it was 97 per cent.

The situation might have been ameliorated considerably if copies of the Student Handbook had been received earlier at the schools, and students and teachers had been able to practise and discuss the sample items in a relaxed atmosphere. The need for prior practice seems most important if the "construction shift" item is to be used in future. Possibly, also, more time should be allowed for this type of exercise.

A number of teachers questioned what the "construction shift" was intended to measure, a question not raised for the "sentence correction" type of item. Evidently appraisers were less impressed than the Advisory Committee had been with the fact that in the "construction shift" exercise the student himself must manipulate sentence elements, a task not called for in any other available multiple-choice item type.

Respondents were asked for an overall evaluation of the difficulty level of the language achievement parts of the test (0--too easy; 5--too difficult).

	13	12A	12G	CAAT	UNIV
too easy	3	1	0	0	9
somewhat easy	11	7	0	3	13
about right	53	51	26	38	61
somewhat difficult	29	31	39	53	17
too difficult	5	10	3	6	6
	----	----	----	----	----
Mean	3.21	3.43	4.10	3.71	2.86

It is interesting to observe here that the teachers of advanced Grade Twelve generally saw the test as close to appropriate in difficulty (despite the general concern, noted above, concerning the "construction shift"), whereas general Grade Twelve teachers saw the test as too difficult. The means of the university and Grade Thirteen responses are approximately equidistant from (3) "About right". The responses suggest that tests of this order, while appropriate for the university-bound students for whom they were originally designed, may be inappropriate for the senior student without university ambitions.

Regarding the suitability of multiple-choice format and alternatives, the following questions were asked in the context of the combination of the language test with (for a sub-sample of students) the essay.

Do you think a test
in multiple-choice
format is is gen-
eral a satisfactory
means of assessing
students' language
competence?

	13	12A	12G	CAAT	UNIV
Yes	28	30	36	24	12
Yes (qualified)	44	45	37	61	76
No	28	25	27	15	12

Where multiple-
choice tests are
used, should they
be supplemented by
other measures?

	13	12A	12G	CAAT	UNIV
Yes	87	90	86	82	96

What importance do
you place upon a
sample of the stu-
dent's writing in
an evaluation of
language competence?

	13	12A	12G	CAAT	UNIV
essential	93	89	79	91	96
important but not essential	7	11	18	9	4

of minimal im- portance and utility	1	0	2	0	0
neither im- portant nor useful	0	0	1	0	0
How do you regard the use of both a multiple-choice test of language achievement and a sample of writing in assessing lan- guage competence?	13	12A	12G	CAAT	UNIV
The multiple-choice test is satisfactory by itself.	2	1	3	4	0
The use of both is important.	69	73	68	80	84
The writing sample is satisfactory by itself.	28	23	26	15	16
Neither is partic- ularly satisfactory.	1	3	3	0	0

If a student's relative standing on a writing test were different from the student's relative standing on a multiple-choice test of the type administered in this study, which would you consider the most valid measure of the student's language competence?

	13	12A	12G	CAAT	UNIV
the score on the writing test	49	53	56	44	56
the score on the multiple-choice test	3	0	2	0	0
a combined score weighted in favour of the writing test	42	43	39	38	40
a combined score weighted in favour of the multiple-choice test	3	2	0	3	0
a combined score giving equal weight to both tests	3	3	3	15	4

A significant percentage responding to these items strongly objected to the use of multiple-choice tests at all as a measure

of language achievement, the higher proportion of strong negatives being found at the secondary panel. As well, a very substantial percentage gave only qualified approval of their use (over 40 per cent secondary, 61 per cent CAAT, and 76 per cent university). Virtually no one regarded a multiple-choice test of language achievement as by itself adequate as a measure of student competence in using language. In fact, a substantial number saw the writing sample as satisfactory by itself (range here was from 15-28 per cent of responses), and an overwhelming proportion of all groups regarded the writing test score as capable of standing alone or, if in combination with the multiple-choice test, with extra weight given to the writing test. The percentages here were: 13--91 per cent; 12A--96 per cent; 12G--95 per cent; CAAT--82 per cent; UNIV--96 per cent.

One rather obvious conclusion should be drawn concerning present or future testing of language competence at the senior level: secondary school, CAAT, and university teachers of English do not think multiple-choice testing valid as the sole indicator of achievement; writing must also be tested. At the same time, it is reasonable to conclude that many teachers would find acceptable a multiple-choice test in combination with writing. One evaluation issue, therefore, that needs to be systematically addressed is a method for increasing the reliability and validity of scoring students' written work.

Respondents' written comments offered a variety of suggestions regarding the multiple-choice format, such as the addition of some test of oral competence or tests of sentence style and structure wherein the student would compose his own sentences. A substantial number of respondents asked that multiple-choice tests provide for short "essay type" answers to questions, though the kinds of questions to be answered in that form went unspecified.

Some criticism was directed against the complicated instructions for the "construction shift" items as noted above; a number responding for the general level student saw the test as

too difficult and "painful" for these students, and rather irrelevant as well. Scattered criticisms were made of testing conditions, the pressure of time and the guessing factor. An infrequent but interesting comment pointed out the total lack of context for sentences to be corrected or improved. Standing alone, they appeared as purposeless exercises, divorced from meaningful communication situations. The absence of context would appear to be a constraint on validity, and a factor to consider in future test construction.

The major specific criticisms, occurring with high frequency, were these:

- (1) The lack of any necessary connection between performance on these exercises and the way in which the student actually uses language in a variety of situations.
- (2) The tendency of these tests to overemphasize artificial, overly subtle, or debatable features of language, with the concurrent tendency to stress awareness of rules rather than performance. Likewise, the tendency to stress errors and error-correction rather than presentation of effective written English.
- (3) In providing answer choices (applicable in this test only to the "sentence correction" items), the failure to test the student's ability to construct his own sentences.

On this last point, a substantial number recommended open-ended items wherein the student would write corrected sentences from scratch rather than be cued by the options.

The above criticisms strongly suggest that the state of the art of multiple-choice language testing is presently unsatisfactory. Teachers are evidently not convinced of the validity of this very indirect measure of language performance nor

are they pleased with the kind of emphasis these tests encourage in approaching the study of language.

2.6 The Essay

As must already be apparent from the previous group of responses, there was a strong consensus that some sample of actual writing, principally the essay, ought to be included in any test of language competence. It was in anticipation of and in agreement with this conviction that the Advisory Committee made sure that at least a sub-sample of Anglophone students being tested for English language competence wrote essays. In the inventory for the Writing Test (the essay), respondents were asked to consider the single mode (expository) chosen for the essay test: its suitability and its importance.

How many students
should be able to
write an acceptable
essay of this type
upon entry to
English courses at
this level? (Post-
secondary inventor-
ies phrased it as
"entry to your
postsecondary in-
stitutions".)

	13	12A	12G	12B	CAAT	UNIV
100%	66	35	24	25	45	80
76%-99%	23	38	16	33	30	16
51%-75%	9	23	33	25	15	4
26%-50%	1	1	7	0	0	0
1%-25%	1	1	7	17	0	0
none	0	0	0	0	0	0

How much emphasis
do you give in your
teaching to the de-
velopment of stu-
dent competence in
this mode of writ-
ing? (For post-
secondary:

"in your teaching
of first year
courses".)

	13	12A	12G	12B	CAAT	CAAT	UNIV
						REG.	REM.
heavy	52	52	32	58	39	73	44
moderately heavy	38	45	46	33	48	7	52
light	2	1	2	0	0	7	0

How many of the stu-
dents who success-
fully complete En-
glish courses at
this level should be
able to write an ac-
ceptable essay of
the type required in
the test? (For post-
secondary: "the
first year English
course you teach".)

	13	12A	12G	12B	CAAT	CAAT	UNIV
						REG.	REM.
100%	77	60	23	17	47	47	88
76%-99%	19	33	44	67	37	20	12
51%-75%	3	6	20	17	16	7	0
26%-50%	1	1	11	0	0	20	0
1%-25%	0	0	1	0	0	0	0
none	0	0	0	0	0	6	0

The Advisory Committee's judgement that this mode of writing is important in Grade Twelve and Grade Thirteen English at all levels was well supported. 90 per cent of the responses for Grade Thirteen indicated moderately heavy to heavy emphasis, 97 per cent for Grade Twelve advanced, 78 per cent for Grade Twelve general, 91 per cent (admittedly of a very small number of responses) for Grade Twelve basic with competence expected by the end of the program at any level by 75 per cent or more of the students according to the large majority of the respondents, and by all those responding for universities. It is interesting also to note that, according to these responses, this mode of writing receives quite heavy emphasis in postsecondary programs.

Are there other modes
of writing which from
the standpoint of gen-
eral literacy are as
important as, or more
important than, the
mode being examined?

	13	12A	12G	12B	CAAT	UNIV
No	50	53	48	--	48	54

Was the present as-
signment at a reas-
onable level of
difficulty for stu-
dents in courses
at this level?

	13	12A	12G	12B	CAAT	UNIV
Yes	94	99	86	79	94	100

Was the restriction
to a single mode
fair to students in
a test of writing
competence?

	13	12A	12G	12B	CAAT	UNIV
Yes	81	78	71	62	66	92

Given the restriction
to a single mode,
what is your opinion
of the range of
topics?

	13	12A	12G	CAAT	UNIV
Good	58	55	47	38	40
Satisfactory	36	40	40	53	48
Unsatisfactory	6	5	13	9	12

Teachers were strongly agreed that the difficulty level was reasonable and that the topics selected and the range of topics were satisfactory. There was some criticism of the restriction to a single mode on the grounds that not all students could demonstrate their best writing when confined to this mode, but, conceding that limitation, it does appear clear that, of the modes that might have been selected, the choice made was the soundest one.

When asked in a further question whether there were other modes which, from the standpoint of general literacy, were as important as, or more important than, the mode selected, between 48 and 58 per cent across the programs and panels represented agreed there were not. Those who disagreed gave a very wide range of responses, including the following: the essay in literature; reports; letters; personal, creative or "free" writing; dialogue

or drama; description; narration; poetry; reviews; and the formal essay.

Of these, narration and description received the most frequent mention but no particular response occurred more than 15 per cent of the time, with two exceptions: for business communications students at the Grade Twelve general and CAAT levels, letter writing was strongly recommended; and for CAAT students generally, informative descriptions of processes and proceedings were frequently suggested.

Teachers were asked to rank in order of importance five general criteria for evaluating the essay, and to suggest any others they considered of high importance. The five criteria were: organization; logic, use of evidence; style (chiefly sentence style); grammar, usage, mechanics; and diction. Some respondents regarded the question as pointless, claiming either that the criteria given were too general or that all were of equal importance. Some, especially among university respondents, gave all five a "1" (high) rating. Consequently, all that comes clear from responses to this item is the strong priority given organization and logical argument over the other general criteria across all grades and levels. It might be noted also that "grammar usage and mechanics" were given priority over "style" at the postsecondary level.

In suggesting additional criteria, many teachers reinforced the importance of organization by specifying it further in such terms as "coherence", "effective introduction and conclusion", etc. A wide variety of other criteria were mentioned but, with one exception, with very low frequency. The one additional criterion which occurred with high frequency (perhaps in 25 to 35 per cent of responses) and in many guises is perhaps best described as "creativity". Evidently, many teachers regard "originality", "vitality", "imaginativeness", etc. as quite as important as more formal and more easily defined criteria.

It should be noted that the essays were in fact scored "holistically" (i.e. by general impression) rather than against a set of formal criteria. The tendency of teachers to give major characteristics approximately equal weight provides some support for the procedure employed.

For the holistic scoring process, scorers were not provided with a set of criteria to apply. However, following a practice run on a small batch of essays, all scorers were asked to identify in rank order the five or so criteria they found themselves applying. The two criteria most frequently noted were organization (almost always in the first or second rank), and logic or development of argument (most frequently in the second rank). "Grammar, usage, mechanics" was the criterion noted third most frequently, and ranked on the average fourth. Creativity and its approximate synonyms also occurred quite frequently, as did diction. In short, the criteria and their ranking as applied by the scorers appear to correlate closely with the views of the teachers as expressed in the inventory.

Following the holistic scoring, a sample of the essays was studied for error counts (grammar, usage, mechanics, etc.) and further characterized from the standpoint of these other more general characteristics.

In various contexts throughout the inventory, suggestions for ways of increasing the validity of measuring language competence were solicited from respondents. Some, though the number was not large (about 10 per cent), recommended increased frequency of testing, an increased battery of tests, or both. A substantial number recommended that oral language be tested in some manner, though there were few specific suggestions. Very few indeed recommended tests of formal grammar. A significant number suggested the précis and the appreciation, no doubt in recall of the external examinations in Grade Thirteen.

The dominant response, however, was the recommendation of more frequent samples of writing in a mixture of modes. This response came forward again in even stronger terms from those making concluding or summary remarks.

Here are some typical comments underlining the importance of the testing of writing:

"A fair test of writing (but) there is a need for a number of samples in various modes over time."

"The writing test is far superior to multiple-choice where there is much guessing, no application, little sense of accomplishment."

"The essay is most valid because it measures organic relationships."

"The writing test is the most effective way of evaluating the student--but problems in subjectivity."

There were a large number of somewhat hostile or negative comments from secondary school teachers, few from postsecondary. Apart from those comments directed against multiple-choice tests on principle and those noting the difficulty of the test battery for general and basic level students particularly, most centred on the timing (the point in the school year) of the tests, the lack of adequate time for preparation on the part of staff and students, the impersonality and apparent lack of direct value of these tests to the students taking them, or insufficient time for students to complete the tests adequately. While this sort of criticism appears fair enough, it is somewhat beside the point of the testing issues to be considered.

On the other hand, there were numerous favourable comments about both the multiple-choice and essay tests with recommendations for improvement. In particular, there appeared to be considerable support for formal testing at the senior level in English.

2.7 Conclusions

Generally, appraisers across all levels offered strong support for the use of multiple-choice tests as measures of both reading and language competence, provided that these tests be supplemented by a sample--or, preferably, samples--of student writing. Appraisers felt students should be given the opportunity to write in more than one mode, but that the "point of view" essay was the most important single mode at the interface level. Further, the principal criteria for evaluating this mode of writing at this level were organization, logic, and "creativity".

Regarding the specifics of multiple-choice language tests, appraisers offered critical comments regarding test conditions, wording of instructions, absence of and emphasis on errors in certain item types and a general lack of opportunity for students to create their own sentences.

3. TECHNICAL ISSUES

3.1 Scoring the Test of Reading Comprehension and Language Achievement (English)

Both forms of this multiple-choice instrument were scored to correct for the effect of guessing. The rule that was employed in this correction was as follows: one mark was awarded for each correct answer, one-fourth of a mark was deducted for each incorrect answer, and all questions for which no answer was given were ignored.

3.2 Scoring the Writing Test

The essays were to be appraised holistically by three independent scorers on a scale of 1 (lowest) to 10 (highest), whole numbers to be used. As can be seen in the "Instructions to Scorers" (presented as Appendix A1A of this technical report), holistic scoring, as described in Britton (1966) and NAEP (1972), is a rapid impression score: the total effect of the essay on the reader. No sub-scores on specific criteria are requested.

Each scorer was asked to read a number of the essays in order to establish use of the scale in a preliminary way and then to make an early adjustment if he found that he was not using the total scale effectively. Our only insistence was that each scorer use every category of the scale: some essays had to be rated "1", some "10" and so on. We did not insist or request that the essays be equally distributed across the scale. Thus essays were appraised in relation to each other rather than against an external criterion.

Consistent with this method, direction to scorers was kept to a minimum (see Appendix A1A). Each scorer was given a copy of the instructions to students, a copy of the essay topics, and a general instruction (#3) to appraise from a positive standpoint and in the context of the mode of writing assigned.

3.3 Recruitment of Scorers

It was established that, inclusive of the 50 essays to be scored by everyone (see below), no scorer should receive more than 300 essays for holistic scoring. Knowing in advance the approximate number of students scheduled to write the essay, we determined that 36 scorers would be required, and we decided that each essay should be scored by one postsecondary English instructor and by two secondary school teachers of senior English. Therefore we sought 12 university/CAAT scorers and 24 secondary school scorers.

As testing was scheduled for late May and the essays had to be collected, xeroxed and packaged before they were sent to the scorers, we calculated that scoring could not occur until the second half of June and we were fearful that good scorers might be scarce at that time of the school year. But with the assistance of the executive of the Ontario Council of Teachers of English, we were able to identify a pool of about 125 potential scorers, most of whom, when contacted by letter, expressed willingness to assist.

The selection of the twelve postsecondary scorers was relatively simple, as there were fewer replies here; we decided to use 6 from CAATs and 6 from universities and followed the principle that one or at most two from each of institutions as widely separated geographically as possible would be enlisted.

In drawing the remaining 24 (secondary school) scorers, we split the respondents into those with more than ten years' secondary school teaching experience and those with ten or fewer years' experience, using twelve from each group. The twelve then were drawn having regard for geographical distribution so that scorers would be widely spread geographically and so that there would be representation from both urban and rural areas.

With the exception of the 50 essays which were assessed by all scorers, each essay was scored by three persons: one instructor in either a CAAT or a university; and two secondary school teachers, one with more than ten years' experience, and one with ten or fewer years' experience at the secondary level.

3.4 Preparation of Scorers

Several weeks prior to their receipt of the Interface essays, scorers were sent fourteen essays of varying quality from Grade Twelve or Thirteen students and from CAAT students, together with an explanation of holistic scoring, the background of the

current project, and instructions for the warm-up exercise using the fourteen essays.

With their scoring sheets for the warm-up exercise, scorers were asked to return a list of the criteria they considered most important in judging this sort of writing, and an indication of the relative priority they attributed to each of those criteria. These data were later used in approaching the appraisal of essays for positive characteristics but were definitely not used in instructing scorers in the use of specific criteria for the holistic scoring of Interface essays.

The chief purposes of this warm-up run were as follows: to familiarize scorers with the holistic scoring method and to provide further direction where needed respecting use of the whole scale; to anticipate questions or problems so that these would not interfere with the rhythm of the scoring of the Interface essays; and to stand in place of a formal training session which, given the spread of scorers around the province, would have been very costly. (Note: our "net" of a score reliability of 0.74, compares very favourably with similar endeavours in holistic scoring with three scorers, and suggests that a formal training session with all scorers gathered in a single location is unnecessary.)

3.5 Main Essay Marking Tasks

The main task of marking some 1600 essays was done following the same basic procedure used in the warm-up exercise. Each marker received a stack of approximately 210 essays. These included the 50 essays that were scored by all markers. The way in which the 50 were chosen and the different stacks of essays were prepared for the scorers is described in the main report (see Chapter Two, Part A, subsection 3.6). The reason for having 50 essays scored by all markers was to implement a special essay scoring procedure. The necessity for this procedure is explained in

Chapter Two, Part A, subsection 4.2; the procedure itself is described in Appendix D2.

3.6 Rescoring the Fifty Essays

We had not anticipated an opportunity to have the 50 essays re-scored, as we feared that intrusion of summer holidays would force many potential scorers to refuse our request to do the initial June scoring if we added the rider that a further scoring in July was entailed. However, circumstances later encouraged us to request the additional scoring. For one thing, response to the recruitment program had been surprisingly enthusiastic; for another, since a substantial number of students who were expected to write the essay did not in fact do so, each scorer had received fewer essays than the 300 he had agreed to score. Hence our request for additional scoring went out, and twenty-eight of the thirty-six scorers expressed a willingness to do the job. (That it was, in fact, a re-scoring job was made clear to the twenty-eight in July, before their task began.)

The re-scoring of the fifty essays provided us with important additional information concerning the reliability of marking. A report on the reliability of the essay scoring is provided later in this technical report (see subsection 3.11).

3.7 Difficulty

The difficulty of items in each form of the Test of Reading Comprehension and Language Achievement (English) was estimated for the population of Anglophone students as follows: the percentage of correct responses to an item was determined separately for each school in the sample and the resulting percentages were averaged over all the schools. The resulting averaged percentages are referred to as indices of difficulty--really they are indices of easiness but conventional usage is followed here. The indices for all the items in both

forms are reported in Table A1.1. Separate indices are provided for SSGD and SSHGD students.

The results reported in Table A1.1 clearly support two conclusions:

- (a) The items in the tests were less difficult for SSHGD students than for SSGD students. This finding is not unexpected, however, because the SSHGD students had had a year more schooling on average than the SSGD students and they comprised a more select group of students: the SSGD group consisted of both general and advanced level students whereas the SSHGD group, by definition, consisted only of advanced level students.
- (b) Form 1 of the test was somewhat easier than Form 2. A base line is needed against which to evaluate these results on difficulty. One such baseline is the figure of 60 per cent. This figure is the midpoint between 20 per cent, the percentage of correct answers that would be expected on a five-option multiple-choice question if responses were entirely random, and 100 per cent. Items with difficulty indices at or below the level of chance performance seem inappropriate for obvious reasons, whereas items with difficulty indices near 100 per cent fail to discriminate among examinees. In tests designed to spread students over the range of scores on the test, items having difficulty indices in the middle of the difficulty range, in this case near 60 per cent, are most valuable.

When the average difficulty of items in each test form is viewed from the foregoing perspective, it would seem that both forms were on the difficult side, although Form 1 came close to being ideal for SSHGD students. In addition, both forms contained items having difficulty indices of less than 20 per cent--items of somewhat questionable value in testing SSGD and SSHGD students.

3.8 Speededness

Both forms of the Test of Reading Comprehension and Language Achievement (English) were divided into three sections. These sections were not separately timed in any strict sense, although students were told at the end of the first 10 minutes of the testing period to go on to the second part of the test if they had not already reached that part, and after 20 minutes they were told to go on to the third part of the test if they had not already reached it. Because of this, it seems reasonable to look for evidence of speeding at three points in the test--the end of Part One, the end of Part Two, and the end of the test.

A test is speeded, by definition, if students do not have enough time to respond to all the questions it contains. Because work rates differ, it is impractical to allow sufficient time for all students to answer every item. One rule of thumb that is sometimes used for judging the speededness of a test is as follows: a test is said to be speeded if less than 100 per cent of the students reach the three-quarter mark in the test and if less than 80 per cent of the students complete the test.

As evidence of speededness one looks for a sign that students have had insufficient time to answer the questions in a test. This means looking at the percentage of students who have failed to respond to an item. An arbitrary distinction is sometimes made, when failures to respond are being considered, between an omitted item and an item that was not reached. An item is said not to have been reached if the student fails to respond to it and to every other item that follows it in the test. Otherwise, the item is declared to have been omitted. Accepting this distinction, the percentage of students who failed to respond because they did not reach an item is the index that should be used to judge the speededness of a test. In the analysis of responses to the Test of Reading Comprehension and Language Achievement (English), the percentage of students who "did not reach" is a useful index of the speededness of the third part of the test. But, because a student always responded to

items appearing later in the test than items in Parts One and Two, it is not possible to compute the percentage of students who did not reach items in these parts of the test. Instead, evidence of speededness in Parts One and Two must be sought in the percentage of students omitting an item. If this percentage increases dramatically over the last few items in Part One or Part Two of the test, then the test was clearly speeded.

The percentage of students who omitted and who did not reach an item in each form of the Test of Reading Comprehension and Language Achievement (English) is reported in Table A1.2. These percentages were computed in the same way as the difficulty indices reported in Table A1.1.

Consider the evidence for Part One of both test forms. The division between Part One and Part Two occurred in each form between items 10 and 11. The build-up of omissions from about item 6 through item 10 is very apparent. This build-up, when assessed using the rule of thumb stated earlier, suggests that Part One of each form was speeded for both groups of students. Because the build-up in omissions is greater and more consistent for Part One of Form 2 than for Part One of Form 1, it seems that this part in Form 2 was more speeded than in Form 1.

There is evidence that Part Two of each form was also speeded. The point of division between Part Two and Part Three occurred in the test between items 21 and 22. The build-up in omissions is even greater for Part Two of each form than for Part One. The limits specified in the rule of thumb for speededness are again exceeded. Clearly, the second part of both forms was highly speeded, and, once again, Form 2 seems to have been more highly speeded than Form 1.

The speededness of the third part of each form can be judged from the "not reached" percentages. These are relatively low regardless of the form being considered. The limits specified in the rule of thumb concerning speededness are either completely satisfied or very nearly so. It would be fair to conclude that

the third part of each test was not speeded to any significant degree.

Reasons for the apparent speededness of the second part of each form may be found in the criticisms of the "construction shift" type of item voiced by many teachers in their test appraisal inventory responses. Teachers themselves found the instructions for this item type very confusing and feared that students would also be confused and have their performances affected detrimentally. Unfortunately as well, some students taking the tests had insufficient time beforehand with the Student Handbook, where the item type was explained and practice items supplied. The evidence of speededness shows that the test, as a measure of language competence, did not work as well as it should have. Care should be taken in future administrations of this kind of test, particularly if the construction shift type of item is involved, that ample opportunity be provided for students and teachers to become familiar beforehand with the test's style and purposes.

3.9 Item Discrimination

The biserial correlation between scores on an item--the item is scored 1 for correct and 0 for wrong, omitted or not reached--and scores on the total test provides a crude index of discrimination. In tests designed to spread students over the range of possible scores on the test, the biserial correlation or index of discrimination for a item should be relatively high, say 0.3 or higher.

Indices of discrimination were computed for each item in each form of the Test of Reading Comprehension and Language Achievement (English) using the responses of two groups of students. These groups were subsamples of the total sample of Anglophone SSGD and SSHGD students tested in the study. Each group took both forms of the test, but one group took them in the order Form 1 - Form 2, and the other took them in the reverse

order. Indices of discrimination are reported for the first form taken by a group. It should be noted as well that both groups were composed of both SSHGD and SSGD students. The two groups were formed at random and because they were drawn from the same pool of subjects, they can be said to be randomly equivalent.

The indices of discrimination are reported in Table A1.3. Note that the indices are above 0.3 for all items except three, two in Form 1 and one in Form 2. In view of these results, it can be concluded that the items in both forms had acceptably high indices of discrimination.

3.10 Distribution and Reliability Statistics

Further information about the two forms of the Test of Reading Comprehension and Language Achievement (English) is provided by the statistics reported in Tables A1.4 and A1.5. These statistics were computed from the test responses of the two groups of students described in the previous subsection on item discrimination. Although both these groups of students took both forms of the test, they took them in different orders. The statistics that are reported are for the form of the test that was taken first.

From the mean scores reported in Table A1.4, it is clear that Form 1 was the easier of the two. Although both forms were somewhat difficult for these students--the means of scores corrected for guessing are less than one-half the number of items in the forms--they were not excessively difficult, as evidenced by the presence of perfect scores on both forms.

An internal consistency measure of reliability and the corresponding standard error of measurement of each form are reported in Table A1.5. These statistics provide information about the stability of the score achieved by an individual student. The reliability coefficients are lower than the level of coefficient that can be achieved with tests of this sort. This

seems to be due in large measure to the fact that the content of the test was heterogeneous. The test was purposely made heterogeneous in order to assess several different components of language achievement, but as a consequence, the correlations among the three parts of the test are only moderately high (see Table A1.5). This has the effect of reducing overall test reliability.

Is the level of reliability so low that it will invalidate the results of this study? The answer to this question is an unequivocal, No! The test forms have an acceptable level of reliability for making comparisons among groups and for use as variables in a regression study. These are the two uses to which the instruments were put in the present work.

3.11 Reliability of Essay Scoring

The plan that was followed in scoring the Writing Test made it possible to estimate a coefficient of reliability and a standard error of measurement for the scores assigned to the essays. Recall that a subset of 50 essays was drawn at random from the total set of essays and that each of the 36 markers of the Writing Test marked the essays in this subset. Three weeks after the full set of essays had been marked, the subset of 50 essays was sent to the markers for rescoring. Eight markers were unable to score the essays a second time, and so in the end two sets of marks on the 50 essays in the subset were obtained from each of 28 scorers.

The procedure that was used to estimate the reliability and the standard error of measurement of the essay scores involves the application of analysis of variance to the scores assigned by the 28 markers to the 50 essays on the two occasions. This procedure is described in detail by Cronbach, Gleser, Nanda and Rajaratnam (1972, pp. 42-44, 86-90, 97-99). It consists of estimating the component of variance associated with each of the factors in the design of the experiment for collecting the scoring data and using

these components to compute an intraclass correlation coefficient and the square root of an error variance. These are respectively the estimates of reliability and the standard error of measurement.

The results of the analysis of variance are summarized in Table A1.6. Note that the largest component of variance, as one would expect, is for essays. This indicates that the largest portion of the variation in essay scores is attributable to differences in the quality of essays. The other relatively large components of variance are for the following sources: markers, reflecting differences among markers in standards; the interaction between essays and markers, reflecting the fact that the score a marker assigns to an essay is not wholly explainable in terms of the extent to which that particular marker is a more (or less) severe marker than the average marker and that particular essay is a more (or less) superior essay than the average essay; and the residual, reflecting error variance and variance due to the interaction among markers, essays and occasions. The estimate of the component of variance for occasions was negative, a result that is impossible in theory, but obviously possible in practice. For this reason, it was set to zero, as recommended by Cronbach *et al.* (1972, p. 57).

The reliability of the essay scoring was estimated from the components of variance reported in Table A1.6. All sources of variance except for essays were treated as sources of error. In addition, it was assumed that each essay would be scored by three different markers and that the final score assigned to an essay would be the average of the three scores it was given by the three scorers. The effect of this assumption was to reduce the contribution to error of the components of variance due to markers and to the interactions between markers and other sources to only one-third of the values reported in Table A1.6. The residual component of variance was similarly reduced to one-third the tabled value. A coefficient of reliability is estimated as the ratio of the component of variance for essays to the sum of the components of variance due to essays and to the other sources (as

reduced in the way described earlier in this paragraph). The estimate of reliability is 0.74. The associated standard error of measurement is the square root of the sum of the components of variance attributable to error; this number is 0.90.

The reliability coefficient and standard error of measurement can be interpreted as follows: the reliability coefficient is an estimate of the correlation that can be expected between the "observed" scores for essays, where the "observed" score for an essay is the mean of the scores assigned on just one occasion by three different markers chosen at random from the population of markers, and the "true" ("universe") scores for the same essays, where the "true" score for an essay is the mean of the scores the essay would receive if it were graded just once by every marker in the population of markers. (This population can be thought of as that large group of persons qualified to mark English essays and from which the individuals who did mark the essays were drawn at random.) The standard error of measurement is interpretable as the standard deviation of the observed scores an essay would receive if it were to be marked by many different groups of three markers, each group chosen at random from the population of markers. If it is assumed that such a distribution of observed scores as this for an essay is approximately normal, then the interval obtained by first adding the standard error of measurement to the observed score for an essay, and then subtracting it from the observed score is that interval within which the true score of the same essay would be expected to fall with probability $2/3$.

Is a coefficient of reliability of 0.74 suitably large, and a standard error of measurement of 0.90 suitably small? The answer to this question depends on the use to which the essay scores are put. Given this standard error of measurement, we should not have great confidence in declaring that the higher scoring student of two students who achieve closely similar scores is the better writer. In other words, in comparing individuals achieving similar scores on this test, the risk of being wrong in judging one or the other to be the better writer would be

relatively high. In the present instance, however, the essay scores were used (i) to make comparisons among different groups of students, and (ii) as predictors of school marks. The obtained reliability of the essay scores is satisfactory for these purposes.

Four additional qualifying remarks are in order.

- (a) The estimate of reliability that has been reported is only an estimate of the reliability of scoring. It reflects nothing about the variation that might be found to exist in essay performance if the same students were asked to write more than one essay. The quality of a student's writing is known to vary as a function of such factors as the mode of the writing assignment, the choice of topics offered, motivation to do well and the student's state of mind. Because of these factors, the correlations that one might expect to observe between the scores achieved by the same students on different essays are almost certainly less than 0.74. And yet it is also true that the more samples of a student's writing that are collected and scored, the better the basis one has for judging his ability. Even though students' scores on two essays do not necessarily correlate highly, both essays together provide a better measure of writing ability than does either one taken alone.
- (b) The evidence on reliability that has been presented is applicable to the case where the observed score assigned to an essay is the simple average of the scores given the essay by three different markers. A simple average was not used in this study; instead, the three scores that an essay received were differentially weighted to form a weighted average (see Appendix D2). This being the case, the results reported here do not strictly apply to the essays as they were finally scored. But these results must be a close approximation

to the true results. (Note that the correlation between the simple averages and weighted averages for the English essays was 0.97, and the variances of both the simple and the weighted averages were identical to the second decimal place.)

- (c) The effect of handwriting on the marking was not controlled by having the essays typed. Even a cursory glance at the essays provides convincing evidence that the quality of penmanship varies greatly from student to student. Just how much the markers were influenced by quality of handwriting is not known, but the contribution of this factor to the unreliability of the marking is almost certain to have been considerable.
- (d) The single number that is reported as the standard error of measurement of the essay marking does not reflect the fact that markers disagreed more over some essays than over others, and that, in consequence, the scores assigned some essays were more variable than the scores assigned others. It is a question of some interest why markers achieve a greater consensus over one piece of work than over another. Perhaps, as James Britton has indicated (personal communication), the originality or unpopularity of ideas contained in an essay, and the informality of the tone of an essay are factors that divide markers and cause scores to vary widely. Subjective impressions formed during the conduct of this study suggest that topic and mode may also be related to variability of marking. Systematic study of the influence of these factors on variability of essay marking is called for to see whether this is an avenue to improved reliability of essay scoring.

3.12 The Essays and the Scale

A. Introduction

While holistic scoring provides the most reliable method of assessing essays in relation to each other, the scale itself is arbitrary; scores provide no characterization of the writing quality beyond "better than..." or "worse than...".

It was considered important, therefore, to characterize the writing across the scale, as far as was reasonable, with reference to such questions as the following:

- (a) What sorts of errors were students making, and with what frequency?
- (b) To what extent could frequency of error be seen as influencing the scores given the essays?
- (c) What positive characteristics (of style, organization, effectiveness of argument, etc.) did the essays exhibit?
- (d) What proportion of the essays could be considered to meet a standard of literacy acceptable for entry to postsecondary institutions?

For the error count and the assessment for postsecondary entrance standard of the essays, the 50 essays scored by all scorers were used, as these had been randomly drawn initially and their scores had proven to be of satisfactory reliability. Thus, although they did not happen to distribute themselves evenly across the scale, they served most reliably to reflect the meaning of scores near those points where their scores fell.

For the appraisal of other writing characteristics, these 50 essays were supplemented by an additional 50 drawn at random from each decile in the total distribution of essay marks in order that we would have a larger sampling of styles and of style variation resulting from choice of topic.

It must be stressed that all observations and generalizations made in the following sections are very tentative, first of all because the sample used was not large, and second because there is a good deal of subjectivity in what a person defines as "literacy", "good style", etc. There is even a surprising amount of subjectivity in what is considered an error, or a major error as distinct from a minor error, regardless of the detail provided in instructions to those counting errors (Britton et al., 1966).

We hope, nevertheless, that these analyses will provide some factual basis for drawing conclusions about the present quality of student writing at the interface, conclusions somewhat different from the general charges, largely unsubstantiated except by the occasional quoted "horrible example", appearing almost daily in the press. It is worth mentioning too that further analysis of the writing, making use of a larger sample, would undoubtedly tell us more about the actual situation, and put into another perspective the voices of mere opinion.

There are two other caveats that must be stressed in considering the performance on the essay in the Interface study as typifying writing performance in Grades Twelve and Thirteen.

The first is that the students in the study were, for practical reasons as described earlier, restricted to a single mode: the expository essay developing a point of view. While the results from the Test Appraisal Inventory strongly confirmed our choice of mode as the best if only one could be chosen, still, it is certainly true that not all students write as effectively in this mode as they do in others; therefore, ideally, they should

have an opportunity to write in at least two modes (Britton et al., 1966, p. 25).

The second caveat is the problem of motivation. The students in the study were "semi-volunteers" for the tests and essays, and pressures of time in May did not give some of them much chance before writing the tests to study the Student Handbook or to ascertain the purposes of the study. Motivation of many was low; there seemed little benefit for the individual in the exercise. That this adversely influenced performance (or non-performance) was evidenced by the number of absentees, by comments from teachers on the Test Appraisal Inventory, and by flippant or cynical remarks appended by some students to their essays. As well, some essays were rather deliberately off topic or beside the point as students chided the tests and the context of the testing. We must therefore be careful not to assume that the writing generally typifies that which would emerge if the students had been highly motivated either intrinsically through personal interest in the topics or extrinsically by examination marks or other rewards.

B. Analysis of Errors

The purposes of the error count were these:

- (a) To obtain a reasonable picture of the sorts of errors present in the students' writing and the frequency of those errors. The error count was a concrete response to the general criticism of student writing at the interface. The public is repeatedly being told by prestigious persons that "most" or "many" secondary school graduates can't write a complete sentence, can't spell, write ungrammatically, are "functional illiterates", and so on. The error count was conducted in an attempt to measure the truth of these charges.

(b) To determine the extent to which error frequency of various sorts appears to influence the scores given essays in holistic scoring. From the Test Appraisal Inventory as well as from independent studies such as Paul Diederich's Measuring Growth in English (1974) and the just-completed The Queen's English (Norman, 1976), it is apparent that satisfactory performance in "grammar and mechanics", variously described, is perceived as an important component of literate writing. Often, from the emphasis this factor receives, it appears to be considered a conditio sine qua non of acceptable writing.

Three sources were examined in preparing for the error count: the Britton et al. study (1966), the National Assessment procedure in their first round of writing assessment (Report # 8, for 1972), and the list of major and minor errors in the instructions to markers of past Grade Thirteen English Composition examinations in Ontario, published repeatedly in S.4C and regularly used by teachers in marking writing in the senior grades.

In general we followed the procedure outlined in the National Assessment, restricting ourselves, however, to 200 running words rather than the 300 used in the NAEP, as our essays frequently did not run to 300 words. We used two "counters" and averaged counts.

With the exception of paragraphing, we used the National Assessment error classifications, but with considerable modification to bring the classifications into close alignment with the types of errors recognized in the Grade Thirteen "departmental" examinations. Error classifications and instructions to the error counters are presented in Appendix A1B of this technical report.

Paragraphing errors were not included in the count for the following four reasons: 200 running words generally do not provide a sufficient number of complete paragraphs; the question of what constitutes a paragraphing "error" as distinct from an inferior or "so-so" paragraph is rather debatable; the National Assessment found seriously defective paragraphs a very infrequent occurrence anyway; and research on the nature of the paragraph (topic sentence, etc.) in current good writing practice reveals much greater variety than standard language texts would imply (see Braddock, 1974).

In the error counting procedure we provided a more elaborate classification of types of error under grammar and diction than is reported; however, as total errors per 200 words for specific types of grammar and diction errors were quite low, we have combined them for ease in assimilating the information.

The main results of interest are presented in Tables A1.7 and A1.8. (Ignore, for the moment, all the numbers in these tables that appear in brackets.) The first of these tables contains means and standard deviations. It should be recalled, as these results are studied, that the number of spelling errors in an essay contributes to the number of errors in conventions for the essay and that the total number of errors for an essay is the sum of the number of errors in conventions and the number of errors in grammar, sentence structure and diction.

An important observation that can be made about the results presented in Table A1.7 is that, for all the error counts, the mean number of errors over the 50 essays is only a very little larger in magnitude than the standard deviation. This state of affairs reflects the fact that the distributions of error counts are positively skewed, with most essays having a number of errors less than the mean and a very few essays having a number of errors very much larger than the mean.

The correlations reported in Table A1.8 are very much in line with intuition. Two points at least merit emphasis.

(a) The global rating of essay quality is negatively related to each of the error counts and the correlations are moderately large. This suggests that the quality of an essay as estimated through the global rating procedure used in this study is associated to an appreciable extent with the number of errors an essay contains. Another way to interpret the correlation of -0.58 between global rating of quality and total number of errors is as follows: given this correlation and the standard deviations of these two variables, the slope of the regression of errors on rating can be computed. This slope is -2.0 , signifying that for every increase in judged essay quality of one mark (on the scale from 1 to 10), a corresponding decrease of approximately two errors can be expected.

(b) The various error counts are positively related, indicating that students prone to make one kind of error (e.g. spelling) are also prone to make other kinds of errors (e.g. grammar, sentence structure, diction).

A difficulty arises in interpreting the results presented in Table A1.8. The distributions of errors are skewed. The danger exists that a small number of extreme values on a variable could be responsible for the presence of correlations as high as those that were observed. This possibility was put to the test by "trimming" the data. The four essays having the largest total number of errors were excluded and the analysis was redone. The results for the reduced set of 46 essays are reported in parentheses in Tables A1.7 and A1.8. Note that the effect of excluding the four essays was to lower the means and the standard deviations of the error counts. The drop in the standard deviations for errors in conventions and total number of errors was especially large. But despite these changes in means and

standard deviations, the pattern of correlations (see the figures in parentheses in Table A1.8) was affected relatively little. In particular, the moderately strong, negative relationship between essay rating and total number of errors was maintained. (The slope of the regression of total errors on essay score after "trimming", is -1.2. This result is somewhat more satisfactory than the slope of -2.0 obtained for all 50 essays because the expected frequency of errors in a essay scored 10 is near 0 for the trimmed regression, not an unrealistic negative quantity as for the untrimmed regression.)

The question of what constitutes a tolerable frequency of error in essays at the senior level, whether these be errors in spelling or major errors in sentence structure, cannot be answered from these results except arbitrarily. A frequency of 6 errors in sentence structure, grammar and diction in 200 running words (the mean for the 11 essays judged to be poorest in the holistic scoring) is likely to be viewed as intolerable, while the overall average frequency of 3.3 may be viewed as acceptable if errors, let us say, in conventions, are few; or, alternatively, a higher frequency of convention errors may be tolerable where other errors are few. This appears to be illustrated in the fourth ranked essay (average rating 8.0) which has the relatively high frequency of 4.5 errors in convention.

On another hand, however, the essay may be of exceptional merit on entirely other grounds so that a higher frequency of errors overall is tolerated, as appears to be the case in the ninth ranked essay (average rating 6.9, number of errors 8.5) and even more dramatically in the essay ranked thirteenth (average rating 6.3, total number of errors 12.0). It is apparent from the less than perfect correlation between error count and average rating, that error frequency is not, by itself, the determinant of essay quality.

It is not safe to go beyond the following fairly obvious generalization: there is a moderately strong tendency for essays receiving lower scores to have a higher frequency of errors, but

it is not consistently the case that low frequency of error is reflected in high scores or vice versa. Obviously, there are many other characteristics of writing that must be considered in assessing "competence" or "acceptability" or "literacy".

One other observation should be made. The overall mean error rate of 7.1 per essay includes all the errors, gross or trivial, made in an essay. Further study of the errors reveals that, on the average, students made 0.8 major errors per essay in sentence structure, and 1.1 grammatical errors, both major and minor. If these are regarded as "serious" errors, then, obviously, the number of "serious" errors per piece of writing was considerably less, on the average, than the total number of errors. A decision as to which figure, 7.1 or 1.9 (obtained as the sum of 0.8 and 1.1), more truly reflects the writing capabilities or the writing incapacities of students requires a more or less arbitrary judgment that each reader must make for himself.

C. Features of Organization and Style

In the foregoing section, our attention was restricted entirely to a negative aspect of student writing performance, frequency of error. It is evident that, though error frequency is one means of characterizing writing, it is by no means the only factor contributing to the overall quality of a piece of writing. There are many other more positive characteristics of writing that must also be considered.

Obviously, in discussing "organization", "effectiveness of argument", "logic", "flair" or "style", we are concerning ourselves with matters of judgement not particularly amenable to quantification. But it does not follow that these aspects of writing should therefore be ignored; otherwise a crude and most unfortunate bias would be built into this report.

In order to provide some useful generalizations concerning the quality of writing as seen from these less well defined standpoints, our sample of 50 essays, for which we had highly reliable scores, was expanded to one hundred, the others being drawn in equal numbers from each decile in the distribution of essay scores. The scores on these additional 50 essays were based on only three scorers' ratings rather than thirty-six; consequently, the original 50 will be normally drawn upon as exemplifying writing at different points of the scale.

Generalization from rather loosely defined characteristics of writing in only a selected sample of 100 essays, or even in a sample as large as our total number of essays (approximately 1600), lacks a good deal in scientific rigour; we admit and regret this, and would recommend a more sophisticated analysis of a much larger sample. Nevertheless, the essays written for the study remain a fund of important information on student writing performance at the interface and can serve as useful baseline data in future studies.

The principal resource drawn upon in this study is a study of the scoring of writing that was done in 1961 for the Educational Testing Service by J. French, S. Carlton and P. Diederich as reported in Diederich (1974, pp. 5-10, 53-58). In that 1961 study 300 essays were scored by sixty readers in six occupational fields by a method resembling our holistic scoring procedure in the present study. Subsequently,

"this table (of scores) was subjected to...'factor analysis' which has the effect of picking out clusters of readers from all over the table who agree within their cluster and disagree with every other cluster to a greater degree than could be attributed to chance. In effect, it determines how many different schools of thought exist among the readers as to what constitutes excellence in

student writing. In this study we found five different schools of thought..."

Diederich (1974, p.6)

Further analysis, by examination of the comments of these clusters of readers, revealed that the scorers in the largest cluster were most influenced by the ideas expressed: richness, soundness, clarity, development, and relevance to the topic and to the writer's purpose. The remaining clusters, in descending order of size, were concerned with usage; sentence structure; punctuation and spelling; organization; diction--the quality of wording or phrasing; and flavour or style--with reference to the personal qualities revealed in the writing.

In anticipation of the present analysis, we checked with two groups of participants in the study--the thirty-six scorers and the teachers who answered the test appraisal inventory--the priorities which concerned them in the writing of students at this level.

Following a trial scoring run, prior to their receipt of the Interface essays, scorers were asked to list the criteria most important to them. Their replies--in order of priority--may be summarized as follows:

(a) Organization

(b) Mechanics: usage, sentence structure, punctuation, etc.

(c) Presentation of Argument: validity, evidence, balance, etc.

(This criterion was in part dictated by the mode assigned: an expository essay presenting a point of view.)

(d) Writer Commitment: involvement, writer's own interest
(seemingly related to Diederich's "flavour")

(e) Style and Diction

In the Test Appraisal Inventory, teachers, secondary and postsecondary, were asked to place in order of priority the following criteria as applied to the mode of writing assigned to the students:

Organization; Logic, use of evidence; Style (chiefly the sentence; Grammar, usage, mechanics; Diction

Responses, when averaged over appraisers, placed Organization strongly in the first position, followed by Logic (the "mode-specific" criterion). Secondary school teachers placed Style third and Grammar, etc. fourth, followed by Diction. Postsecondary teachers reversed the order of Style and Grammar. Though many additional criteria were suggested, most fell loosely under one or other of the foregoing classifications.

A further verification of the significance of Paul Diederich's "factors" is found in the former Grade Thirteen departmental examinations' criteria for the "Middle First":

The middle first is excellent in content and style. Some of its merits are comprehensive and intimate knowledge of subject matter (IDEAS), freshness of thought (FLAVOUR), skill in arrangement (ORGANIZATION), command of sentence structure, and consciousness of word values (DICTION).

S4C - Underlining and bracketed remarks
are ours.

Finally, and independently of the present study, the Queen's University report on student literacy (Norman, 1976) has recently been made available. There (pp. 72-73) eleven "basic components

of literacy" are suggested. Four of the eleven "components", ranked as #5, #6, #7 and #10, concern effective presentation of argument: logic, detachment or objectivity, respect for facts, effective generalization. Others which strongly echo the priorities that appear to be established from our searches are (rank is Norman's):

2. The ability to write an assignment that is well organized...

4. The ability to write grammatically, in a style which is free of all but minor errors of spelling, punctuation, grammar and vocabulary, and which makes proper use of English idiom.

1. The ability to write sentences which follow logically and meaningfully from one another, and paragraphs which are coherent and unified. (Appearing to combine style (re sentence structure) and organization again (re paragraph))

No. 3 (fluent; clear, concise writing) and No. 8 (immunity to slang jargon, and colloquialisms) appear to be concerned primarily with diction.

The substance of these excursions into the realm of writing quality is that there is rather considerable agreement concerning the positive attributes of literate writing and concerning those attributes most relevant to postsecondary success. Further, Diederich's original classification, if not necessarily his order of priority, is validated elsewhere, and in our case with one important modification.

The modification is this: that as the mode assigned specifically required the development of an argument or point of view, Diederich's first priority, "ideas", should be characterized more specifically as being concerned with point of

view expressed, and the effectiveness of the logic and evidence employed to substantiate it.

The sample of 100 essays therefore was examined from the following standpoints, making flexible use of a Diederich-type analytic scale (Diederich, op. cit., p. 54) with reference to the positive characterizations of writing as described. The classifications concerning conventions and mechanics were omitted, as they had been separately examined in the error count.

The following criteria (elaborated, with variations, from Diederich, pp. 55-57) were employed:

1. ORGANIZATION (from Diederich, unmodified)

High: The paper starts at a good point, has a sense of movement, gets somewhere, and then stops. The paper has an underlying plan that the reader can follow; he is never in doubt as to where he is or where he is going. Sometimes there is a little twist near the end that makes the paper come out in a way that the reader does not expect, but it seems quite logical. Main points are treated at greatest length or with greatest emphasis, others in proportion to their importance.

Middle: The organization of this paper is standard and conventional. There is usually a one-paragraph introduction, three main points each treated in one paragraph, and a conclusion that often seems tacked on or forced. Some trivial points are treated in greater detail than important points, and there is usually some dead wood that might better be cut out.

Low: This paper starts anywhere and never gets anywhere. The main points are not clearly separated from one another, and they come in a random order--as though the student had not given any thought to what he intended to say before he started to write. The paper seems to start in one direction, then another, then another, until the reader is lost.

2. EFFECTIVENESS OF ARGUMENT OR VIEWPOINT

High: The writer provides a balanced argument or viewpoint, treating the subject with an appropriate detachment and some originality, and the argument develops to an appropriate conclusion based on the evidence. Judicious use is made of examples and other forms of evidence, and generalizations are supported. The argument of viewpoint develops logically. The writer reveals commitment to this position.

Middle: The writer tends to rely too much on unsubstantiated generalizations and general clichés of thought. The viewpoint expressed may be rather one-sided without adequate regard for alternative positions. Transitions between stages of the argument may not always be clear. The student may evince little commitment to what he is saying and betray a low interest in the subject.

Low: Generalizations and thought clichés abound with little regard for substantive evidence. The argument may be heavily biased or, perhaps worse, the writer may attempt to present all sides without ever taking a firm position and the essay rambles to no conclusion in particular. There is little awareness of logical development and the arguments themselves are tired.

3. STYLE (SENTENCE STRUCTURE)

High: Sentences are well varied in relation to purpose. Less important ideas are effectively subordinated and there is clear evidence that such devices as parallelism are used to good effect. Sentences are varied in length and made economical in expression through ellipsis and tautness of phrase.

Middle: Though some variety in sentence structure is evident, the variety is not particularly well related to effect. The writer appears unaware of the effective use of stylistic and rhetorical features of language. There is limpness of style with occasional minor errors.

Low: No awareness of style or variety in sentence structure is evident, and errors, major and minor, do occur. Sentences tend to ramble to no purpose and expression is woolly or redundant.

4. STYLE (DICTION) (from Diederich, unmodified)

High: The writer uses a sprinkling of uncommon words in an uncommon setting. He shows an interest in words and in putting them together in slightly unusual ways. Some of his experiments with words may not quite come off, but this is such a promising trait in a young writer that a few mistakes may be forgiven. For the most part, he uses words correctly, but he also uses them with imagination.

Middle: The writer is addicted to tired old phrases and hackneyed expressions. If you left a blank in one of his sentences, almost anyone could guess what word he would use at that point. He does not stop to think how to say something; he just says it in the same way as everyone else. A writer may also get a middle rating

on this quality if he overdoes his experiments with uncommon words; if he always uses a big word when a little word would serve his purpose better.

Low: The writer uses words so carelessly and inexactly that he gets far too many wrong. These are not intentional experiments with words in which failure may be forgiven; they represent groping for words and using them without regard to their fitness. A paper written in a childish vocabulary may also get a low rating on this quality, even if no word is clearly wrong.

5. FLAVOUR (from Diederich, unmodified)

High: The writing sounds like a person, not a committee. The writer seems quite sincere and candid, and he writes about something he knows, often from personal experience. You could not mistake this writing for the writing of anyone else. Although the writer may assume different roles in different papers, he does not put on airs. He is brave enough to reveal himself just as he is.

Middle: The writer usually tries to appear better or wiser than he really is. He tends to write lofty sentiments and broad generalities. He does not put in the little homely details that show that he knows what he is talking about. His writing tries to sound impressive. Sometimes it is impersonal and correct but colourless, without personal feeling or imagination.

Low: The writer reveals himself well enough but without meaning to. His thoughts and feelings are those of an uneducated person who does not realize how bad they sound. His way of expressing himself differs from standard English, but it is not his personal style; it is the way uneducated people talk in his neighborhood.

Sometimes the unconscious revelation is so touching that we are tempted to rate it high on flavour, but it deserves a high rating only if the effect is intended.

No generalizations based upon an examination of approximately 100 of some 1600 essays can be pressed too far. There are, even within the sample studied, exceptions to most of our generalizations. Nevertheless, some interesting patterns do emerge, sufficient, we hope, to lead to some useful reflection on the character of student writing in Grades Twelve and Thirteen.

Just as 100 essays drawn from across the range cannot be said wholly to typify writing at different points of the range, so, too, it must be remembered that these essays represent work of students ranging quite widely in their present level of schooling and their plans for the future. In addition to Grade Thirteen students (many anticipating university), Grade Twelve students at both general and advanced levels, preparing for work, for further education at CAAIs, or for another year of secondary school before university--all contributed to the essay sample. The exemplar essays have been selected, therefore, to illustrate some of the characteristics of writing and their presence (in degree) or absence in the sample across the range, not as good or bad examples of writing nor as a basis for saying "That's how (all) senior secondary students write."

After a brisk reading of the essays, a selection was made of those that fell at particular points along the scale. These essays were carefully examined and scored according to the criteria described above, using a 1 - 5 scale (1 - low; 5 - high) for each criterion. The first purpose was to determine whether and at what points along the scale essays began to meet individual criteria adequately. Secondly, we wished to see whether some criteria were met by essays scored low as well as high, with little pattern. When scores on the Diederich criteria were compared with essay marks, we found, as expected, that scores on specific criteria tended to rise as the scores on the holistic scale rose; however a number of irregularities in the pattern made it clear that strength in certain

characteristics, notably quality of argument and flavour, were by no means the exclusive preserve of those writers scoring near the top of the scale. As well, the rate of improvement as related to the holistic scale was different for different criteria, as noted in the following comments.

(a) Organization: Essays at the bottom of the scale generally exhibited very poor organization, often with poor and infrequent paragraph division. Conclusions, if reached, tended to be very limp and sometimes essays trailed off altogether.

However, in all the higher parts of the range examined, essays, whatever other weaknesses they had, did exhibit adequate to good control of paragraphing and evidence of reasonable planning. Though there was a failure in some to give due weight to the more important aspects of the argument, they were reasonably coordinated and developed towards a conclusion.

(b) Argument: Quality and force of argument with appropriate illustrations and balanced treatment proved to be much more unevenly distributed. A few of the otherwise weakest essays did show some strength in this regard, though most of the essays that with some consistency could be praised for quality of argument were found in the top portion of the score range.

The general weaknesses exhibited included reliance on superficial generalizations and lack of adequate evidence in support of statements. Some writers failed to discriminate between less and more important aspects of the cases they were making and tended to introduce redundancies or irrelevancies.

(c) Style--Sentence Structure: Serious errors in sentence structure were infrequent right across the range (see preceding "error count" report); on the other hand, attention to sentence style appeared to be generally lacking in essays scored below 6.0. Though there was some variety in sentence length, order and complexity, it appeared rather random, without consciousness of effect.

Use of parallelism and antithesis, which might be expected frequently in the presentation of an argument or viewpoint, appeared very rarely except in the essays rated near the top. In these last essays, there was evidence of tautness and control; in the less distinguished essays, sentences tended to ramble and insufficient use was made of means of subordinating less important ideas.

(d) Style--Diction: Diederich's description of the dull average, "The writer is addicted to tired old phrases and hackneyed expressions," fairly characterizes the writing in a high proportion of the essays examined. The search for the exact word, with a few notable exceptions, simply was not evident. It may be speculated that students were insufficiently motivated to put forward their best efforts; nevertheless, the lack of colour and vigour in diction and imagery is discouraging. The most popular topics (#1 on Violence and Censorship and #4 on Course Options--Secondary) particularly appeared to encourage dullness. Pretentiousness of diction was rare (and almost welcome); those students who wanted to show off did so in other ways.

(e) Flavour: The same topics (#1 and #4), though sometimes argued strongly, had generally a staleness about them.

"Flavour" was not entirely concentrated in the superior essays. Some middle-range essays, undistinguished in other respects, came alive through the evidence of personal commitment on the part of the writer, while some of the essays in the top of the range, and correct in most respects, lacked originality and freshness. While flavour was more commonly found in the best-rated essays than in those with lower ratings, no general statement about the distribution of flavour across the range is possible.

The absence of flavour in so many essays may be a reflection of the testing situation, in which motivation was low and students could expect no writer-audience interaction. It would be unwise, therefore, to conclude that most senior secondary students are without the ability to produce flavourful prose. In fact, the most interesting and enjoyable examples of writing sometimes came from

students who refused to take the test seriously and so entered into a sort of devil-may-care dialogue with the anonymous examiner. The freshness and breeziness of style, while perhaps a little too much of a good thing (as in a following example), produced an appealing personal ingredient otherwise generally lacking. That this occurred with some frequency underlines the importance of developing "real" writing situations in order to measure more effectively student writing ability.

EXEMPLAR ESSAYS

These, as noted earlier, are not meant to typify the student writing but to illustrate some of the writing qualities referred to in the criteria so that the trends can be better understood. Comments on errors in spelling, grammar, etc. have been excluded, as error counts have been treated separately.

FIRST EXAMPLE

(SCORE 3.7)

TOPIC: Secondary schools should not demand that students take particular courses or pre-requisites; students should be free to choose courses of interest to them.

The debate of whether students should have a choice by courses they can take in high school, can have two sides. On the pro side, there would be most students who are about to enter or have just entered the secondary level of education. The reason for this is quite simple, people want as much freedom as they can get. I know, several years ago as a student in grade eight, the idea of freedom of choice was quite appealing. But today, I can see how easily it could have been for me to make a serious error in course choices. This occurred to me only after it was too late to make any alterations. Fortunately, though, I did make the right choices, and therefore, I can continue my

education as I wish. But other people I know have not been so fortunate. They made the wrong choices when they first entered high school, and today they cannot follow the course of education they wish. The main reason for this occurrence is that students are made to make this decision when they are quite young and do not fully realize the importance of the decision.

On the con side of the argument are the older people or people who have finished their education. These people on the whole would like to see the courses pre-set and compulsory to all students. In this way, they make the assumption that this will result in better educated students to enter university. This is a certain degree can be a valid point. But there is one fallacy to the argument, students on the whole will do better in subjects they enjoy taking than in subjects that do not particularly appeal to them. I know personally that subjects I enjoy taking average as much as twenty percent higher than subjects I don't particularly enjoy. This I imagine can be said for all students. The result therefore is that if you make students take many subjects that they do not enjoy, then they will not work as hard at them and their marks will generally be lower than if they had chosen the course themselves.

At this point, it may appear that I am sitting on a fence and not making any alternative methods. But I feel that a compromise can be made between the pro and the cons. I feel that some courses should be compulsory. These subjects are English, mathematics, History and science. English is an important pre-requisite since almost all universities require English before they will accept graduates. In addition, English is a subject that makes the backbone of a person's intelligence. If a person has many good ideas and thoughts, he can be considered intelligent, but if he can't express his ideas in an

intelecual fasion, than he may not appear to be an intelecual.

Mathematics is an important subject becaus for graduates to get into university under various cources in sciencses and engenering and medicine, he may need mathematic cources from secondary schools.

History is a necessary course for students to take. Of all the cources you take at high school, history has one of the most practical uses in every day life. If a person ever gets involved in a discusion, the knolage of historical events may be very useful in defending a point.

Science is also an important course that should be taken, at least in the level one grade, because it gives the student a better understanding of the world around him.

I feel that all the cources I have stated above should be taute at the advanced phase level until the end of the level two year. In this way students are much more aware of their own goals they want to take, whether it being university or colages of applied arts and technology. After the level two year, I think that a greater freedom of choice of cources should be given with english and mathamatics being compulsory.

Therefore, if a compromise between the two sides is drawn with the outline above, the level of education will be bettered without the expen(illegible) of the freedom of the student being taken away.

COMMENTS ON FIRST EXAMPLE

This essay, though replete with mechanical errors, illustrates the point that essays relatively low on the scale did show some

effectiveness in organization and paragraphing. It follows a "Here's one side; now here's the other" pattern, followed by development of the writer's own position in some detail and with reference to several subjects the student considers important. And even if the conclusion is limp, it does provide a summary statement as the student steps back from particulars to a final generalization.

No doubt paragraphing could be more effective; a better design might have been to deal with specific subject areas in fewer paragraphs, but there is at least a sense of "one idea--one paragraph".

The tendency to weak generalizations is illustrated by this essay: "On the con side of the argument are the older people...(who) would like to see the courses pre-set and compulsory to all students," as if there were clearly two camps of the human race, entering Grade 9-ers and "the older people", the one group holding one position universally, and the other, the other. What the student refers to as the fallacy of the latter position (that "I", and likely most other students, do well in courses I like) ignores the possibility that students tend to "like" the courses they do well in; i.e. have the capacity for. On the positive side, the student does attempt to be specific in reference to his own experiences and in noting specific subjects. A reasonable sequence of reasoning can be followed and there is balance in the argument.

In style, both sentence structure and diction, the essay is wholly undistinguished, characterized by a lack of terseness and little consciousness of word values. For example, the writer adds an explanatory sentence, "The main reason for this occurrence is that students are made to make this decision when they are quite young..." rather than attaching the explanation to the previous sentence with a linking phrase such as "chiefly because".

The essay has some flavour: one senses a commitment to the position taken and evidence of deliberateness in its development. The personal note is frequent and desirable; without it the essay would be extremely dull.

SECOND EXAMPLE
(SCORE 5.5)

TOPIC: Should scenes of violence in books, in films, and on television be censored?

It is hard to say whether censorship should exist at all or to what degree. If everyone was of sound-mind and basically a good character we wouldn't have to worry about it. Evidently however even all us Canadians don't qualify for the sound-mind bit. Unfortunately also not all the ones deserving a rubber room have one either so were left with a lot of good people a few harmless crazy ones and a few who aren't so harmless.

I've read a lot of books and enjoyed most of them. I'm just trying to think of them without their crime festured plots. They'd probably be really dull and boring for most people. Maybe it would prohibit one idea of the perfect crime but Canadian illiteracy % would likely skyrocket.

Last year, several schools throughout Canada and the United States were plagued with shooting incidences. A mentally disturbed student would enter a school and open-fire in the halls and classrooms. Many people were killed and everyone seriously frightened.

Undoubtedly, the tremendous amount of coverage on the topic and the general out roar from society against this particular crime somewhat encouraged these

students. It really makes you wonder whether were not all to blame. At the beginning of each new television season we witness more and more crime. Our need for gory deaths and drastic amounts of blood never seems to be fulfilled. Sometimes changing the channel doesn't help much either. Our taste for good "shows" is becoming warped and contorted--we like crime!--

However, I'll admit we're humane. Sure enough after the above mentionned killings new laws prohibiting certain guns and rifles were introduced. How well they are enforced and by whom is up to us. We really can't expect too much if we think well you know--my son can have a rifle--he's got it all upstairs!

Canadian crime statistics have increased over the past few years along with the population. Provincially Ontario came out number one again and not by a small percentage.

If you were to ask a father what crime he considered the worse, he'd likely tell you, rape. thinking back tell September, I can recall hearing about or seeing at least twenty shows involving rapes. Is it so odd then that the number of rapes is also ascending? Whether it be good old-fashioned Dr. Welby, Medical Centre, Policewoman or movies, we cannot escape the present trend. We'd all likely vote rape in our top five most dreaded crimes for women and we'd disown a raper son.

It use to be, a women had to prove her innocence when she pressed rape charges. Her history was gone over and if possible she was made to look like a real tramp. The raper was presented to be the all-american boy. Recently however, a new bill was passed whereby a woman's history would not be revealed when dealing with this sort of crime. Instead the pendulum swings the

other way and the raper must try to prove his innocence. More of this should be shown on television. It might serve to discourage would-be rapers if they thought their chances of getting caught were a lot higher.

I'm just not content to say well were not as bad as the States!

Presently Judy LaMarsh is heading a committee in charge of censoring crime on T.V. I think it is a worthwhile cause if done properly. Right now, I find Canadian T.V. censorship too lax. For example "Deliverance" was on T.V. not too long ago, the next day I saw one little boy grab the ear of another kid and told him to squeal like a pig!

Everyone knows that even little kids don't go to bed right after dinner so shortly after T.V. family hour doesn't really make a difference in the age of the audience.

I find that many movies shown at theatres which are restricted have less to offer in the way of crime as do many T.V. shows.

I agree censorship should exist but to certain degrees.

COMMENTS ON SECOND EXAMPLE

There is an organizational thread running through this essay, though the writer's view of the topic does not become especially clear until late in the essay, because he (she) jumps out early into a number of sociological dimensions, presumably effects of excessive TV violence, before coming back to state the argument or case. There is, then, a plan, though one that could be made more effective.

Here, however, we do see improvement in the effectiveness of argument, though the essay remains somewhat disjointed and some of the assumed causal connections are rather far-fetched (just as they often are in magazine or newspaper articles on the censorship of violence). He is reasonably concrete in his illustrations: note especially, paragraph ten. And among his reflections emerges the possibility that it is not our television that is to blame, but ourselves--the society and what it values. There are insights that prevent him from making blanket generalizations or suggesting simplistic solutions.

There appears to be a genuine attempt to vary sentences for effect, perhaps an over-reliance on exclamation marks. Diction is uneven in quality; sometimes words are chosen with precision, at other times the choice is overly colloquial and there are actual errors. (Or is "outroar" an interesting coinage?)

Though there are many matters to fault in the essay, the writer has compelled the reader's interest with some touches of humour (even if rather trite) and with his directness. Through the writing we catch something of the writer as a person, a feature rather infrequent across the sample of essays examined. The essay does indicate that "flavour" is not a quality of writing exclusive to the best essays.

THIRD EXAMPLE (SCORE 6.0)

TOPIC: What are the values of the Olympic Games, not only for the competitors, but especially for the countries they represent and for the world?

In this essay I will be discussing the values in the Olympic Games, not only for the competitors, but especially for the countries they represent and for the world.

The year 1976 is an important one for Canada. This year, in Montreal, Canada has the honour of holding the summer Olympics which will commence in July. The Olympic Games have had excellent results in their money raising campaigns through the Olympic lottery. This is one source of money for the construction of the buildings many will see this summer. The second source of money, unfortunately comes from the Montreal tax-payer's pocket.

The Olympic's value in a country is the union it brings. Every year either in summer or winter the games that are held in a chosen country unite all countries of the world. Even if there are four athletes present to participate in games there is no limit to the amount of persons that should represent their country.

The Olympics this year in Canada will bring the awareness of how important sports are for Canadians. Millions of Canadians are not fit, hopefully the Games will motivate them to take up one of the numerous sports offered.

The Olympic Games also bring in tourists from all the corners of the world. This not only provides wealth for the country, particularly the city hosting the games but also provides an image of beauty for a country. Once the visitors come to a country that shows warmth and friendship as seen in Canada there will be a want for return or a recommendation to their friends.

For those Canadians who do not attend the Olympic Games, they will be able to see the buildings in Montreal whenever they wish. The buildings in Montreal: the veladrome, the centre arena, and the Olympic Village will remain there, in memory of the honour, beauty and union that the Olympics once offered in Canada.

The Olympics for Canada brings a union within the country. In recent years there has been a major problem of bilingualism. This is heard in the province of Quebec. They feel Quebec's major language should be French which has faded away in the last couple of years. The Games will provide a unity for the east and west side of Canada.

The value the competitors take in the Olympic Games is the most exciting experience in their life. The feeling of representing their country is seen all over their face. Even if the athletes do not gain a medal they are not sore losers but feel they gave it a good try and besides everything else it was worthwhile. The athletes of each country get to communicate and learn about other cultures of their competitor. There is friendship between two countries shown by the athletes.

The athletes show the importance of a particular sport in their country. This is shown by their training, skills, and a medal usually. For example: In West Germany, sports play a major role. After work every individual has a particular activity they take up. This shows the importance of fitness in that country. Athletes work very hard for long hours day after day. They eat certain foods, and not too much entertainment is allowed on weekdays. Almost everything is under strict control. These results lead to the other country's awareness of a higher aim for the next year. Either winners or losers all competitors rate the same to each other. They know due to experience the hardship one must go through to enter the Olympic Games.

Therefore we may conclude that it is through the competitors and the country as a whole that bring out the values of the Olympic Games.

COMMENTS ON THIRD EXAMPLE

This essay, though reasonably adequate in organization, illustrates the general blandness of the writing in the middle of the range. Viewed from the standpoint of errors, mechanical and other, this essay appears to be more effective than it really is.

The naïveté of the argument and lack of substantive support are obvious at once; views of the Olympics as a cure for bilingual problems, an expression of world friendship authored by the host country, and an event with real impact on sports and fitness all seem rather futile (though the reader must remember that the essay was written before the Games). The writer has "covered" the topic, but thinly, and with little sense of direction or purpose. The concluding sentence is "limp"--just a sign-off.

There is little awareness of sentence style (note the choppiness of the many sentences in the ninth paragraph), and diction and imagery remain wholly abstract.

The blandness of effect is explained and reinforced by the absence of flavour; there is no sense of personal commitment or interest. The one touch of humour (last sentence of paragraph two) stands out only because nothing else in the essay does.

While many commentators on student writing at the interface have been greatly concerned about error frequency, the frequency of such apathetic writing perhaps should be a matter of even greater concern. Doubtless the test circumstances were a significant contributing factor in that they did little to stimulate writer commitment; nevertheless, the dryness of so many essays suggests that more might be done in schools to encourage commitment and vigour in writing. The student with a stake in his ideas and with a richer sense of audience is the one who is likely to write purposefully and effectively.

FOURTH EXAMPLE

(SCORE 6.8)

TOPIC: Should scenes of violence in books, in films, and on television be censored?

Have you read *Stranger in a Strange Land* by Robert A. Heinlein. If not, get up off your ass and go buy a copy. On second thought stay there and I'll tell you the pertinent point. (How's that for alliteration?) Heinlein says that man is the only creature that laughs and that mankind laughs "because it hurts". This a very emotional world with an abundance of pain, both physical and mental. Is pain an emotion? A feeling? No, it is pain; intangible but ever present. Pain is a guillotine over your head that is ready to drop at any instant. Pain hovers above you at all times and each one of us becomes changed in some way due to our efforts to avoid its presence.

Does violence warp a child's mind? Does he become "strange" because of his efforts to avoid or perhaps even seek the pain he has been exposed to? I was weaned on violence and look how well I turned out. Or did I? Now turn around and look at the occupants of the world's prisons and asylums. And graves. Most of those people weren't born that way: They were changed, warped by society. Somewhere along the line they were twisted by their environment.

It is time to try a little preventative medicine. We can never abolish violence but we can try control the market we sell it to. Go ahead and censor the violence on television during the hours that children watch. Censor that idiot box while you're at it. But remember that if you initiate your censorship program you will put a lot of industries out of business. Certainly the cop shown would survive: there will always be an adult

market for such trash. But what about the cartoons and comic books. Even Walt Disney films have violent scenes. Those comic books are really incredible: "Triple A rating--No sex, no drugs, no foul language", just picture of three mice going through a meat grinder and a cat swallowing a bundle of dynamite. When was the last time you watched Bugs Bunny? Did you see him shoot Elmer in the face with a shotgun? Funny eh? Have ever laughed at a friend who just fell on some slippery ice? Don't deny it, you thought it was hilarious. You laughed because it hurts. The Three Stooges built a comedy empire on pain and violence. Think of the last good joke you heard. Chances are that if the joke wasn't dirty it's meaning was actually rather sad. If you eliminate violence you will eliminate a great deal of laughter too.

I guess you realize that I haven't taken a firm stand on this issue. I believe that we would solve a lot of emotional problems, clear a lot of mental wards, if we were able to stop teaching violence to impressionable youngsters. But I can't be sure that those children would be able to cope with the violence they are sure to meet in the "jungle". I'm not sure that I could stand to see people living on half-rations of laughter either.

I'm laughing now because it sure as hell hurts.

COMMENTS ON FOURTH EXAMPLE

This essay is included to illustrate the observation made in the introductory comments that some of the most interesting and flavourful writing was exhibited where the student refused to take the test too seriously. The writer-reader relationship changes dramatically from a dreary "writer to examiner" style, illustrated in the previous essay, to more of a "writer to his peers" style.

Probably because this student has broken out of the mold, he provoked a wide variety of responses from the scorers who rated the essay all the way from 3 to 10 on the scale. Apart from the mechanical faults, evidently some scorers reacted strongly and negatively to abrupt shifts into a language register generally colloquial and of questionable appropriateness. Other scorers likely over-reacted in the opposite direction, simply out of relief at the gasp of fresh air this writer provided. No one can say the essay is weak on flavour, even if the flavour is too strong for many.

Leaving aside "register", there is an unevenness in the writing and argument. The quasi-philosophical reflection on pain seems partly irrelevant. And flippancy is frequently indistinguishable from mere carelessness. There is a logic to the essay but the personal intrusions and illustrations, even if germane, are introduced in a way that partly obscures the direction of thought.

On the other hand, many of the jabs are good ones. Are not the observations on contents of comic books sharp and fair? And the writer is not the first observer of the human scene to note the interplay of pain and laughter. On a second reading, it is clear that this student operates well in mingling the concrete and specific with the universal.

Diction, barring the "bottom of the vernacular" descents, is generally precise if not outstanding, and certainly the writer uses sentence style to create effects.

A quick and vigorous mind is at work here, revealing itself in somewhat heterodox ways. And even if these are perceived as somewhat extreme, the essay suggests that if a more effective stimulus were provided, one that placed writer and intended reader in a more intimate relationship, rather different and much more interesting qualities of writing might emerge.

FIFTH EXAMPLE

(SCORE 8.0)

TOPIC: Does art imitate life, or does life imitate art?

A piece of art that imitates life is not a true work of art. A person that lives his life imitating art is not truly alive. Art and life are not carbon copies of each other, but rather reactions to experiences.

When an author sets out to write a book he doesn't regard events. That is the reporters job. Instead, the author writes about reactions to the events, from inside him rather than from outside. The artist who paints a picture does not try to imitate the scene. A camera will serve that purpose. The artist will attempt to record his reaction to the scene presented. The ballet dancer can perform the steps perfectly but unless his or her own interpretation of the steps is added, the performance is dead, simply an imitation of life. We have examples of the limitations. Every corner book store has that one shelf of books Mother forbade you to look at. These are imitators and can hardly be called "art." The difference between true art and an imitation of life is the difference between DaVinci and Playboy; between Robert Wagner and Alice Cooper; or between D. H. Lawrence and the man who sits in his attic apartment spewing out 5 books a week.

Life, like art, is a series of spontaneous reactions. By attempting to imitate anything, especially art, we inhibit these reactions and become only half a person, the biological part. What makes us human is not only our physical selves but also our spiritual beings. Just as our body would reject an imitation, any attempt to supply imitations to the spirit would also be rejected. Living a life that attempts to imitate art would be like gazing at a photograph. Oh, the pictures

are very nice but everyone is a bit too stiff and unnatural, the smiles, a trifle too wide.

As a carbon copy is fainter than the original, an attempt to imitate life as art would never be true. Life and art are similar as they are both reactions but reactions can never be recorded and replayed. Here today, gone tomorrow is what makes life and art what they are.

COMMENTS ON FIFTH EXAMPLE

This topic, not a very popular one, produced some very interesting essays. Those most interesting were the ones where the students attempted, as in the present example, to carry mental explorations somewhat further than maturity and knowledge warranted: truly an "essai" or a "try". Generally scorers treated even relatively unsuccessful "tries" with generosity.

This essay illustrates well the quality of essays placed near the top of the score scale. Though the essay is short it is characterized by tautness of language and organization. Statements of position, even if somewhat erroneous in fact, are put forward bluntly and unambiguously. This directness and absence of padding gives a freshness of flavour so frequently lacking in other essays and the choice of language is generally appropriate to the subject matter.

Argument is well developed by use of contrast (life: art; artist: reporter; body: spirit) and use of illustration.

That the student is reaching beyond himself is apparent from his conclusion; he is out of his depth and does not know quite how to reach shore. The reader must, however, admire the swim.

In this essay, we see a more mature effort to relate sentence structure to purpose. If the writer had learned the semi-colon, he might have used ellipsis and antithesis to better

effect. However there is a good rhythmic sense and usually variety in structure relates to effect intended.

SIXTH EXAMPLE

(SCORE 9.2)

TOPIC: A competitive environment develops strength of character.

Today's society is to a great extent a competitive society. Early in life we are taught the importance of "getting ahead" and "striving to be better than the neighbour." Throughout our school system, "success" is measured directly according to one's ability to achieve "more" than one's classmate. When looking for a job, one is forced to compete in order to make an impressive impression. Indeed, in a capitalistic society such as our own, the "competitive environment" is a predominant factor. Some people see no harm in this: philosopher Thomas Hobbes insisted it was "in man's nature" to compete.

Is this so? Or does society force man to be competitive? And if so, does this competitive environment develop strength of character?

Because of the continuous indoctrination of competitiveness throughout one's early life, it is difficult to lose the competitive nature. As a result, men are forever "looking out for themselves"; trying to better themselves; trying to "get ahead." In a capitalist system, the ultimate result is the exploitation of man by man, of class by class. Individuals strive for personal, material gain. Instead of developing strength of character and humane values, men become obsessed with making money. Because our

society is competitive, and because it places a high value on material wealth, it naturally follows that men strive to further personal wealth. The "profit motive" overrides the importance of humane values.

As a result, those who are unable to "compete" live often in worlds of mediocrity or even poverty, while those who are successful competitors live in affluence and are looked up to by society. These "successful competitors" feel little or no obligation to enhance the living conditions of their fellow men. The competitive environment in which they have been brought up has conditioned them into believing that they owe nothing to their "competitors". There is no consideration for the needs of those who are less fortunate.

Instead of building a strong character--one which has consideration for others--the competitive environment conditions man to be narrow-minded and concerned only with self gain. Moreover, the competitive environment creates hardships for those who are unable to compete. Because of our society's overemphasis of the competitive environment, many children go hungry and many elderly citizens are unable to live with dignity. Hopefully, our society will recognize this fact soon, and transform it from a competitive society to a society concerned with brotherhood.

COMMENTS ON SIXTH EXAMPLE

This essay represents the top of the scale as available to us in the sampled essays. It requires very little comment.

Like the previous example, it is tautly constructed; there are no waste words or irrelevant ideas. Plan and direction are always clear.

In argument the student articulates his position clearly; however, as in most of the essays, supportive evidence or illustration is not available. In this essay, the logic of the case is there and the care for development has the reader nodding rapidly in agreement rather than looking for evidence. He would have to look hard, for the essay is built on generalizations about "people", "man" or "men", "some people", and "our society". The student provides a definition of "a strong character" which is rather noble, but it is not a definition that would win universal agreement and it is not supported.

Flavour is good simply because the student feels strongly and expresses himself effectively. The reader, whether he agrees with the position or not, knows whom he is listening to.

Sentences are varied effectively in relation to purpose. The use of a series of questions in a short paragraph is a very effective means of shifting from description to argument. The periodic sentence with the "lead-in" by means of subordination, if somewhat over-worked, is generally effective, especially in development of the argument sequence.

The level of diction, if not exceptional, is generally suited to the topic, but marred by the student's tendency to italicize (by quotation marks) too often.

As in many of the essays, the conclusion is limp. In this instance, as in the previous example, the exaltedness of theme leaves the student floundering for an adequate conclusion.

D. Postsecondary Appraisal of the Essays

The fifty essays scored by all 36 scorers and for which, as a consequence, we have highly reliable scores were examined by instructors experienced with first year students at four Colleges of Applied Arts and Technology and four universities.

The fifty essays were randomly drawn in approximately equal numbers (26 Grade Thirteen, 24 Grade Twelve) from both grades involved in the Interface survey, and included the essays of Grade Twelve students either going on to Grade Thirteen, dropping out, or going on to postsecondary education, and of Grade Thirteen students going on to postsecondary education or dropping out.

The essays were arranged by student code number rather than rank order and instructors were asked simply to read the essays and classify each as

- (A) Acceptable quality of writing for entry to program at this institution.
- (R) Remedial attention appears to be required, but some qualities of the writing appear promising for success.
- (X) Quality of writing not acceptable, even if some remedial attention were provided, for likely success in programs at this institution.

A copy of their instructions is included in Appendix A1C.

Our contact person in the English Department at each of the CAATs and universities selected was asked to find one person experienced in teaching first year students in each of (i) the English Department, (ii) a humanities or social science area, (iii) a technical, business, or natural science area (one CAAT contact person found two judges from the English Department and only one from another department). The contact person made the selection and distributed the essays and instructions, providing any further explanations that seemed necessary. The resulting breakdown of teachers was as follows:

	<u>University</u>	<u>CAAT</u>
English Department	4	5
Other Department	8	7

for a total of 24 teachers.

We were interested in determining the percentage of essays written by students in the different SSGD and SSHGD groups (see Chapter Three for a definition of the various groups) that would be placed in the acceptable, remediable and unacceptable categories by different types of instructors. As well, we were interested in whether there was any difference between English Department members' judgements and the judgements of instructors in other departments. The point here is that a great many students do not elect to major in English at postsecondary institutions: "literacy" acceptable for one department may not be the same, or nearly the same, for all.

A word of caution is in order: Our classifications "A, R, X" are very loose ones and the number of instructors responding is quite small. One presumably could find a wide variation of opinion about the essays within a single department of an institution and a wide one as well among universities or among CAATs. The present analysis is inevitably a very tentative one: it is suggestive but by no means conclusive.

Analysis

For each postsecondary teacher, a separate analysis was made of the relationship between the teacher's categorization of the essays and the score derived from the 36 original scorers. This analysis consisted of two probit regressions, using the essay score as the independent factor and either of the following two dichotomies as the "response":

(i) Acceptable versus Requiring Remediation and Not Acceptable

(ii) Acceptable and Requiring Remediation versus Not Acceptable.

The probit analyses provide functions relating score level to probability of response classification, and these functions depend on the response level (hardness/softness) of the classifier and also on the correlations between the classifiers' responses and the original scores.

In order to see how the postsecondary teachers would have classified typical essays from the five basic groups of secondary students--that is, SSHGD-POSTSEC, SSHGD-OTHER, SSGD-SEC, SSGD-POSTSEC, SSGD-OTHER--each probit function was combined with the distribution function of essay scores for each group. (Recall that a distribution function was a quadratic fit to the logistic transformation of the cumulative proportion of students in a group who were assigned selected essay scores or less. These functions and the probit functions were integrated on a 100-point grid.) The result was an estimate for each teacher and each secondary student group of the percentage of students in each group writing essays classified as Acceptable, as Requiring Remediation, and as Not Acceptable.

Finally, the percentages for the four groups of postsecondary teachers--University English, University other, CAAT English, CAAT other--were aggregated: the mid means, which are the arithmetic means of the middle halves of the distributions, were calculated.

Results

Table A1.9 gives the percentage of essays for each combination of postsecondary teacher group and secondary student group estimated by means of the procedure described above to fall into each category. It is clear that the university teachers are generally

harsher in their judgements: for the student group comprised predominantly of students going on to university (SSHGD-POSTSEC), university instructors would find only about half the essays acceptable. It is interesting to note that the university English teachers tend to declare the other essays--those that do not fall into the Acceptable category--to be Not Acceptable, while the other university teachers more often favour remediation.

The CAAT teachers find a somewhat higher percentage of the essays acceptable, though of the group comprised mostly of students planning to enter a CAAT (Group SSGD-POSTSEC), only 31 per cent wrote essays considered acceptable by CAAT English teachers, and 42 per cent wrote essays considered acceptable by other CAAT teachers. Remediation is suggested for 55 per cent and 48 per cent of SSGD-POSTSEC students, depending on the type of CAAT teacher.

One other result seen in Table A1.9 is the difference between SSHGD and SSGD groups. This is further confirmation of group differences described and discussed in Chapter Three of the report.

The method of analysis used here has led to results having the appearance of great consistency. The reader should be warned again that the sample of teachers was small and cannot be described as a probability sample of any know population. Moreover, the variation in judgements was substantial. For example, the variation among the 12 university teachers in the percentage of SSHGD-POSTSEC essays judged to be acceptable was from 5 per cent to 79 per cent. The variation among the 12 CAAT instructors in the percentage of SSGD-POSTSEC essays judged to be acceptable was from 14 per cent to 53 per cent. For all these reasons, it would be foolhardy to make strong statements based on these results about the acceptability of essays written by different groups of secondary students as judged from the point of view of university and CAAT instructors.

REFERENCES

- Braddock, R., Lloyd-Jones, R., and Schoer, L. Research in written compositions. Urbana, Illinois: National Council of Teachers of English, 1963.
- Britton, J.N., Martin, N.C., and Rosen, H. Multiple marking of English compositions. Examinations Bulletins 16, 1966.
- Coffman, W.E., and Kurfman, D.A. A comparison of two methods of reading essay examinations. American Educational Research Journal 5, 1968, pp. 99-107.
- Cronbach, L.J., Gleser, G.C., Nanda, H., and Rajaratnam, N. The dependability of behavioral measurements. Toronto: Wiley, 1972.
- Diederich, P.B. Measuring Growth in English. Urbana, Illinois: National Council of Teachers of English, 1974.
- Diederich, P.B., French, J.W., and Carlton, S.T. Factors in judgments of writing ability. Research Bulletin 15, 1961. (This report is very difficult to obtain, but an account of the Carlton study is given in Measuring Growth in English--see above.)
- Godshalk, F.E., Swineford, F., and Coffman, W.E. The measurement of writing ability. Research Monograph 6, 1966.
- Hoyt, C. Test reliability estimated by analysis of variance. Psychometrika 6, 1941, pp. 153-160.
- National Assessment of Educational Progress. Report No. 8: Writing Mechanics. Washington, D.C.: U.S. Government Printing Office, 1972.

APPENDIX A1A
SCORING OF ESSAYS

INSTRUCTIONS TO SCORERS

1. Following these instructions is the page from the examination booklet listing the eight topics and giving the instructions to the student. The last page is a reproduction of the instructions from the outside of the booklet. Note first that occasionally you may find an unruled page included in the essay where the student, owing to time constraints, has not had time to re-copy. Please note as well that the student has been told to make his corrections by neatly stroking and writing above the excised material as we did not expect there to be time to re-copy. In your impression scoring, therefore, do not penalize for this inevitable untidiness.

2. You are asked to give your mark on your impression of the whole performance. Subtotals for vocabulary, style, etc. are not to be used. You are asked to make up your mind quickly, keeping to a rate of 20-25 scripts per hour.

Errors in spelling, mechanics and grammar will be separately scored for a large sample of the essays at a later stage as part of our effort to analyze and describe the writing of secondary school graduates. You should not therefore concern yourself with any tally of these.

3. Look for excellences rather than penalize deficiencies, rewarding the writer involved enough to write in a direct and expressive way, and detached enough to show a consistent point of view.

The student was invited to agree, disagree, or to take an intermediate position with respect to the point of view stated or implied by the topic; consequently if the student takes a point of view or approach at variance with the topic statement the essay is not to be considered "off topic".

4. Give a mark of 1 through 10 (1 low: 10 high) to each composition. Please use the whole scale. Scripts you consider the best are to receive 10 and scripts you consider the worst, 1. The numbers 1 - 10 are wholly arbitrary; please do not think in terms of percentages, passes, or failures.

It is not necessary that you seek to place equal numbers of essays in each category of 1 through 10, though you should use all ten.

5. The zero is not part of the scale. Record a zero only where the student has not attempted to write a composition or where there are only fragments of notes.
6. Record scores in the box in the upper right hand corner of page 1 following the student number. For the scores 1 to 9 use the double digit 01, 02, ..., 09. Please use a pen with a color different from black (e.g. red).
7. Return total package in the original box using the self-adhesive return address label provided. Use registered mail. Please endeavour to have the box in the mail by Monday, June 28.

SUGGESTIONS FOR ESTABLISHING CATEGORIES AND DEVELOPING A RHYTHM

In order to establish a standard in your mind read 20-25 scripts selected at random from the whole set you have received before you begin to mark. Place these scripts in 10 different piles representing the scale 1 to 10. After that try to make an individual judgement on each piece of work according to this standard, adding the script to its appropriate pile.

You may find, as these compositions are the product of both general and advanced level twelves as well as thirteens, that the range of ability represented is more extensive than your sample has indicated. If so, revise your classifications to cover the range, re-sorting those essays already piled. At the point where your standard is firmly established in your mind, pause to add the scores to those essays already appraised. Then continue scoring and piling. Try to work quickly, maintaining a rate of 20 or more scripts per hour.

Please be alert to the problem of boredom, especially where you happen to run into a string of essays on the same subject. It is well to plan breaks when you are becoming somewhat saturated.

INFORMATION

Each script is being scored by three people and their perceptions, as represented by the score given, will be combined to produce the raw score for the essay.

APPENDIX A1B
THE ESSAY ERROR COUNT

EXTRACTS FROM INSTRUCTIONS TO COUNTERS

Familiarize yourself with the error classifications, a few of which may appear somewhat arbitrary, and perform a trial run on a few of the essays to develop an efficient rhythm. Check the errors noted against the classification chart to ensure that the instructions are being followed consistently.

The classification of errors on the chart is inevitably limited and arbitrary. You will have to make a number of "judgment calls", and a few errors that perhaps bother you especially may not fit any of the classes. For the latter, with regrets or curses, ignore them. For the others, sooner or later you will have to be arbitrary, so, for your own sanity, make it sooner. I found the greatest difficulty in classifying types of serious sentence errors, for, once a student dives into a construction he cannot handle, there are multiple consequences.

The division allows for

F - Fragments and unanalyzable sentences

R - Run on and run together sentences

K - Seriously awkward, contorted sentences

and then for some common classes of error related usually only to a portion of the sentence:

MM - Misplaced modifiers other than verbals and verbal phrases

MV - Misplaced, mis-related, or dangling participles, gerunds, infinitives

S - Faulty subordination

II - Faulty parallelism or faulty ellipsis affecting a large sub-unit of meaning in the sentence

Very serious faults are to be classified as well as you can under the first group.

I have attempted to cover most contingencies in the explanation of the code, but other ambiguous situations will inevitably emerge. Choose what you see as the most appropriate classification and carry on.

This error count is being carried out simultaneously by two people, and matters should balance out by averaging results.

The essays appear in order #1-#50. This order is unrelated to the scores received.

Procedure

1. The first 200 words, omitting the title, have been marked off for you.

Your error count is to include the first 200 words only, with the following exceptions:

If the 200th word is the last word in a sentence, the end punctuation is to be included in the count.

If the 200th word is not the last word in a sentence, the whole sentence should be examined for errors falling under the sentence structure classification only, not including punctuation except where it contributes to the sentence structure error.

Sample: 200th word It was a wierd\$ and wonderfull
 game and which the Dodgers lost
 didn't they.

Count mis-spelling of "weird" but not the mis-spelling of "wonderful".

Count the "and which" error as a major error (F) in sentence structure.

Don't count the failure to include a comma after "lost" or the failure to end with a "?"

2. Mark the count for each error classification (Sp, Pc, etc.) in the appropriate box beside the appropriate essay number, and give the subtotals in the boxes provided headed "TOT". Put the total error figure in the box to the left marked "GRAND TOTAL". The count on fictional essay #99 in the first set of boxes should clarify the procedure.

Do not include in your count errors that do not approximately fit the categories, and make rapid

judgement calls where there is some ambiguity about error and category provided.

3. For the first 5 or 6 use a pencil to mark the errors on the essay, referring frequently to the classification of errors material, and record counts on your rough copy of the chart.

Then go back over these essays to determine consistency with which you are using the classification making such revisions in classification as seem necessary. Use a bright coloured pen or marking pencil to score over your penciled scoring and to revise your rough counts.

4. Continue with the remainder.

CODE AND EXPLANATION OF ERRORS
ERROR COUNTS

CODE

(1) CONVENTIONS

Sp

Spelling

1. Count a mis-spelling of the same word only once.
2. Accept both American and British spellings: penalise-penalize, theater-theatre, etc.
3. Ignore obvious slips and failure to break at the syllable when a word carries to the next line.
4. Classify all apostrophe errors as spelling errors.
5. Errors in principal parts of verbs should be classified as errors in verbs, not spelling errors.
6. Abbreviations properly spelled are to be accepted. This is a question of style, not of spelling.

Pc

Punctuation

1. The use of comma or no punctuation between principal clauses that should be separate sentences falls under "The Sentence" below. Do not classify it under punctuation.
2. A comma before "and" in a series is usually optional. Don't penalize.
3. For introductory prepositional phrases and short subordinate clauses, usage is variable. Don't penalize unless the sense of the sentence is adversely affected.
4. Failure to put commas both before and after a parenthetic expression should be penalized as one error:

Example: My father who is bald loves me.

1 error

My father who is bald, loves me.

1 error

5. Penalize only once the failure to place the appropriate punctuation marks inside rather than outside quotation marks. (Where the student has carelessly placed quotation marks almost exactly

above the punctuation mark, give him the benefit of the doubt.)

6. Failure to put quotation marks both before and after the words directly quoted--1 error, not 2.
7. Ignore obvious slips on periods. Actually, the photocopying process may not have picked up a light dot, and a punctuation mark at the end of a line may have been missed in the photocopying.
8. Treat the dash generously. Mark as a punctuation error only if clearly wrong, not if it represents a tendency to excessive informality in style.
9. Accept titles either underlined or in quotation marks (not both at once.)

C

Capitalization

1. Count an error involving the same word only once.
2. Geographic locations, book titles, and other proper names may contain one or more words. Count only one capitalization error on each occasion, as:

"british columbia"

"Rime Of The Ancient Mariner"

3. The conventions of English permit only language subjects to be capitalized, as in English, French, Spanish. Count an error in capitalization of school subjects ONCE ONLY. (Otherwise students selecting topic 4 where the problem is frequent are being disadvantaged.)

Note: This convention does not appear to extend to school departments as Department of History, Mathematics Department.

- (2) SENTENCE STRUCTURE (Classify all very seriously faulty sentences under F, R or K.)

F

Sentence Fragments

and sentences that defy analysis.

R

Run-On Sentences

Main ideas tagged together with "and", main ideas separated only by a comma or by no punctuation--"The Comma Splice".

K Awkward
Seriously disjointed, contorted or incoherent sentences, not mere lack of smoothness

MM Misplaced Modifier
Word, phrase, or clause (other than verbals, noted under MV.)

MV Verbals
(infinitives, gerunds, participles) dangling, misplaced, misrelated. Do not mark split infinitives.

S Faulty Subordination
including orphan "which" clauses, "when", "where", and "because" clauses improperly used as nouns.

e.g. I saw you were not at home which is why I broke in. (Orphan "which")

The reason I came is because I was lonely.

II Faulty Parallelism
affecting the sentence. Include faulty ellipsis under parallelism errors.

(3) VERBS

S-V Subject-verb agreement
Be tolerant of collective nouns which, depending on context, may express a singular or plural idea.

T Faulty tense or tense sequence

PP Principal parts of verbs
as in "saw", "has seen", or "have lain" as distinct from "have laid". Do not classify these under "Spelling".

(4) PRONOUNS (Include pronominal adjectives)

A Antecedent
Lack of an antecedent (omitting orphan "which", described under subordination). Ambiguity owing to doubt concerning antecedent. Shift from one person/number to another in the sequence. "Don't penalize the shift back."

C Case of Pronoun

The objective case following a copula verb is acceptable: "It was me."

(5) PARTS OF SPEECH

PS Misuse of a part of speech

e.g., prepositions as conjunctions, adjectives as adverbs

(6) DICTION

D Diction

word obviously inappropriate in meaning or in accepted usage, as "amount" for "number", "between" for "among", etc., and "anywheres", "irregardless", etc. sorts of error (what the Grade 13 examiners formerly called "crudities").

Rd Redundancy

Obvious redundancy where the student uses in close succession words or phrases obviously of the same meaning. These will be situations in which the sense obviously escapes the student.

Rg Register

Grossly inappropriate choice of register of language.

A consistently informal breeziness of style is not to be penalized, nor is the occasional choice of an "inappropriate" word, evidently chosen for effect, even if you are not pleased with the effect.

DO NOT

PENALIZE:

-Sentence fragments written deliberately for effect (even if you are unimpressed by the effect.)

-Split infinitive. Frequently the abuse is tolerated and sometimes it is necessary to simplify style.

-"Due to" when misused. The misuse is now virtually the custom.

-"Shall" and "will" confusions. The distinction, if there ever was one in normal usage, is dropping out of the language.

-Contractions such as "I'll", "hasn't", "don't",

etc. One's tolerance or intolerance of these are matters of acceptable style, not of error.

-Objective case following the copula verb: "It was me".

-"Who" rather than "whom" at the beginning of a sentence as in "Who(m) were you going with?" Established usage.

-Minor faults in parallelism not affecting the sentence structure.

FURTHER DIRECTIONS:

When a serious error in structure has another minor error (e.g., of punctuation) associated with it, treat the event as a single error, the more serious.

e.g., "Besides this serious threat science has been the cause of ..."

The error lies in the lack of grammatical relationship of the opening phrase to the remainder and should be so classified. Ignore the failure to provide a comma after the opening phrase. This would be effectively to penalize the same problem twice.

If, however, "serious" had been mis-spelled, you would count both the sentence error and the spelling error, as they are unrelated.

Treat the issue of hyphens generously.

There is such a wide variety of opinion on what constitutes a compound word that it would be unfair to penalize the student for your predilections, however strongly held.

Ignore errors in the title.

APPENDIX A1C

INSTRUCTION SHEET FOR ESSAY APPRAISALS

Background: You are likely aware that, this past spring, a study was made of student achievement at Grades 12-13 and of the programmes, secondary and postsecondary, in a large number of educational institutions in Ontario. In fact, you may have already been directly involved through responses to questionnaires on the tests (OISE) or on programme (Queen's).

In English, we tested all students in the sample in reading comprehension and language, and, for a subsample, we also provided an essay to be written in 75 minutes. The instructions to students and topics are enclosed. The essays have each now been holistically scored by at least three markers on a 1-10 scale, so we now know how the essays range across the scale. Fifty of the essays, those enclosed, have been scored by all 36 markers on the marking team so that we have been able with these to designate their position on the scale with confidence.

Intent of Present Appraisal: We now need to characterize the scale in several ways, one of which is to obtain information from the post-secondary level from a variety of departments as to which essays represent a satisfactory level of writing ability for success at post-secondary institutions. Certainly, a single sample of writing is hardly an adequate basis for classification; nevertheless, your appraisal will assist us to describe the state of affairs more exactly than the 1-10 scale by itself could do.

The essays are not in order by score, nor is the small sample representative of all the essays received. They are numbered from 1-50 as on your return sheet. The other number 1-8 is not a score; it simply indicates the topic chosen.

Request: We are asking, through the contact person delivering this package, an instructor from each of three different departments or divisions within your institution to give a brisk reading to these fifty essays and classify them on the return sheet under the headings provided there. Recognizing that different characteristics of "literacy" may be required for success in different sorts of postsecondary programmes, we are soliciting your impressions from the standpoint of programmes represented in your department or division rather than in general.

We have left it to the contact person in your institution as to whom to approach; if for some reason you are unable to carry out this assignment for us, please assist him/her in

passing the package on to a colleague and in explaining the circumstances.

A rapid reading of the essays is all that is required. You do not have to mark anything, make notes in the margin, etc. We hope you will not find it a time-consuming task.

Once your return sheet has been filled in, please destroy the essays. Send back the return sheet only.

TABLE A1.1

Percentages of Examinees Responding Correctly
to Each Item in Both Forms of the Test of
Reading Comprehension and Language
Achievement (English)

Item No.	Difficulty Indices			
	Form 1		Form 2	
	<u>SSGD</u>	<u>SSHGD</u>	<u>SSGD</u>	<u>SSHGD</u>
1	86	91	61	70
2	59	77	72	83
3	57	79	45	61
4	25	38	46	60
5	7	14	20	27
6	66	79	58	73
7	54	71	41	53
8	40	54	35	50
9	25	47	18	22
10	18	29	13	19
11	76	82	64	77
12	65	71	62	76
13	46	64	75	87
14	48	57	48	61
15	39	54	48	57
16	32	47	40	53
17	31	43	41	50
18	35	42	19	27
19	22	30	30	36
20	25	34	6	10
21	25	32	10	14
22	72	77	65	74
23	64	80	58	66
24	67	81	64	72
25	83	88	68	77
26	33	39	62	65
27	54	67	33	46
28	34	42	44	53
29	67	76	26	33
30	19	33	41	43
31	64	77	16	21
32	36	53	19	25
33	72	79	44	58
34	26	32	46	58
35	65	67	58	69
36	40	38	50	59
Mean	46.5	57.3	42.9	52.4
S.D.	20.5	20.5	18.8	20.8

TABLE A1.2

Percentage of Examinees Omitting and Not Reaching Items in Both Forms of the Test of Reading Comprehension and Language Achievement (English)

Item No	Form 1				Form 2			
	SSGD		SSHGD		SSGD		SSHGD	
	%Omitted	%Not Reached	%Omitted	%Not Reached	%Omitted	%Not Reached	%Omitted	%Not Reached
1	0	-	0	-	1	-	1	-
2	4	-	2	-	1	-	1	-
3	3	-	2	-	5	-	4	-
4	4	-	4	-	4	-	3	-
5	4	-	4	-	12	-	10	-
6	6	-	4	-	12	-	7	-
7	12	-	8	-	14	-	11	-
8	25	-	19	-	20	-	15	-
9	31	-	23	-	22	-	19	-
10	26	-	19	-	31	-	31	-
11	0	-	1	-	4	-	5	-
12	13	-	12	-	3	-	2	-
13	8	-	5	-	3	-	2	-
14	5	-	4	-	15	-	16	-
15	8	-	5	-	12	-	11	-
16	23	-	21	-	15	-	11	-
17	12	-	10	-	24	-	23	-
18	14	-	12	-	28	-	27	-
19	24	-	21	-	32	-	33	-
20	31	-	30	-	39	-	47	-

TABLE A1.2 (continued)

Item No	Form 1			Form 2		
	SSGD %Omitted	%Not Reached	SSHGD %Omitted	SSGD %Omitted	%Not Reached	SSHGD %Not Reached
21	30	-	28	42	-	46
22	0	-	0	1	-	1
23	0	-	0	0	-	1
24	1	-	0	1	-	2
25	0	-	0	0	-	0
26	2	-	2	0	-	1
27	0	-	1	0	-	1
28	0	-	1	2	-	2
29	0	-	0	6	-	7
30	0	-	0	1	-	1
31	1	-	1	2	1	4
32	1	-	1	0	1	0
33	4	1	3	2	4	2
34	0	2	1	1	6	1
35	0	5	0	0	8	0
36	0	9	0	0	13	0
						12

TABLE A1.3

Indices of Discrimination for Items in Both Forms
of the Test of Reading Comprehension and
Language Achievement (English)

<u>Item No</u>	<u>Form 1</u> <u>(N=1612)^a</u>	<u>Form 2</u> <u>(N=1619)^b</u>
1	46	49
2	53	55
3	71	51
4	49	44
5	38	34
6	55	59
7	63	55
8	54	62
9	69	38
10	64	58
11	44	64
12	38	65
13	57	67
14	54	57
15	61	55
16	61	52
17	60	53
18	41	63
19	52	54
20	55	48
21	56	56
22	41	52
23	65	42
24	63	29
25	51	43
26	39	31
27	37	50
28	29	48
29	36	38
30	57	36
31	54	36
32	65	48
33	49	54
34	42	52
35	29	52
36	33	43
Mean	51	50
S.D.	11	10

TABLE A1.3 (continued)

Note: Decimal points have been omitted from the table.

^aThis group of SSGD and SSHGD students took both forms of the test in order Form 1--Form 2.

^bThis group of SSGD and SSHGD students took both forms of the test in order Form 2--Form 1.

TABLE A1.4

Statistics Describing Both Forms of the
Test of Reading Comprehension and Language
Achievement (English)

	<u>Form 1</u>	<u>Form 2</u>
Number of Examinees	1612 ^a	1619 ^b
Number of Items	36	36
Mean Score	16.49	14.45
Standard Deviations	7.45	7.35
Highest Score	36.00	36.00
Lowest Score	-5.25 ^c	-4.75 ^c
Reliability ^d	0.83	0.82
Standard Error of Measurement ^d	3.05	3.09

^aThese examinees took both forms of the test in the order Form 1--Form 2.

^bThese examinees took both forms of the test in the order Form 2--Form 1.

^cNegative scores arise as a consequence of the application of a correction for guessing; in this case, each wrong answer was scored $(-1/4)$, whereas each correct answer was scored $(+1)$.

^dComputed by applying a formula due to Hoyt (1941).

TABLE A1.5

Correlations Among Scores on the Subtests and
Total Test of Both Forms of the Test of Reading
Comprehension and Language Achievement (English)

<u>Test Part</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
Reading Comprehension Subtest		55	51	80
First Language Achievement Subtest	56		55	83
Second Language Achievement Subtest	56	55		86
Total Test	82	84	86	

Note: Decimal points have been omitted. Correlations below the diagonal are for Form 1; correlations above the diagonal are for Form 2. The sample size on which the correlations below the diagonal are based was 1612; for the correlations above the diagonal, the sample size was 1619.

TABLE A1.6

Summary of the Analysis of Variance of Scores
Assigned to 50 of the English Writing Tests by
Each of 28 Markers on Each of Two Occasions

Source of Variance	Degrees of Freedom	Mean Square	Estimate of	
			Variance Component	
Essays	49	133.20		2.32
Markers	27	32.94		0.25
Occasions	1	0.00		- ^a
Essays x Markers	1323	2.60		0.59
Essays x Occasions	49	2.06		0.02
Markers x Occasions	27	7.23		0.12
Residual	1279 ^b	1.41		1.41

^aThis was a negative number, subsequently treated as zero.

^bFor 50 essays, 28 markers and 2 occasions, this number should be 1323, making a total of 2799 d.f. Unfortunately, 44 scores of the total of 2800 were missing, either because of administrative error--occasionally an essay was not included in the bundle sent to a marker--or because of marker error--occasionally an essay was returned unscored. In the analysis of variance, each missing score was estimated by the sum of the grand mean and the main effects for the essay, the marker and the occasion for which the score was missing.

TABLE A1.7

Means and Standard Deviations of Average Global
Rating and Error Counts for 50 Essays
(Errors per 200 Words)

<u>Variable</u>	<u>Mean</u>	<u>S.D.</u>
Average Global Rating of Quality	5.5 (5.7)	1.5 (1.5)
Number of Spelling Errors	1.8 (1.5)	1.7 (1.4)
Number of Errors in Conventions (Spelling, Punctuation, Capitalization)	3.8 (3.1)	3.4 (1.9)
Number of Errors in Grammar, Sentence Structure, Diction	3.3 (2.8)	3.0 (2.1)
Total Number of Errors	7.1 (6.0)	5.1 (3.3)

Note: The results for a subset of 46 essays are given in parentheses; see text for explanation.

TABLE A1.8
Intercorrelations Among Average Global Rating and
Error Counts for 50 Essays

<u>Variable</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
Average Global Rating of Quality	1.00	(-0.43)	(-0.42)	(-0.45)	(-0.53)
Number of Spelling Errors	-0.52	1.00	(0.87)	(0.36)	(0.74)
Number of Errors in Conventions	-0.43	0.78	1.00	(0.33)	(0.79)
Number of Errors in Grammar, Sentence Structure, Diction	-0.50	0.47	0.28	1.00	(0.83)
Total Number of Errors	-0.58	0.79	0.83	0.79	1.00

Note: The correlations for a subset of 46 essays are reported in parentheses above the main diagonal; see text for explanation.

TABLE A1.9
Estimated Classification Percentages

	<u>University</u>		<u>CAAT</u>	
	<u>English</u>	<u>Other</u>	<u>English</u>	<u>Other</u>
Per Cent Acceptable				
SSHGD-POSTSEC	52	47	62	66
SSHGD-OTHER	43	40	54	61
SSGD-SEC	34	33	44	54
SSGD-POSTSEC	22	24	31	42
SSGD-OTHER	18	21	26	35
Per Cent Requiring Remediation				
SSHGD-POSTSEC	29	41	34	32
SSHGD-OTHER	34	47	42	37
SSGD-SEC	33	50	48	41
SSGD-POSTSEC	32	52	55	48
SSGD-OTHER	29	50	56	53
Per Cent Not Acceptable				
SSHGD-POSTSEC	19	12	4	2
SSHGD-OTHER	23	13	4	2
SSGD-SEC	33	17	8	5
SSGD-POSTSEC	46	24	14	10
SSGD-OTHER	53	29	18	12

APPENDIX A2

TECHNICAL REPORT ON THE TESTS OF FRENCH FOR FRANCOPHONES

This report is divided into three main sections. The first section consists of a description of the Test de compréhension en lecture et de connaissance de la langue (français) and the Test de composition écrite. The results of an appraisal of these tests are reported in the second section. In the third and last section, information about the technical properties of the tests is given.

1. DESCRIPTION OF TEST CONTENT

The two forms of the Test de compréhension en lecture et de connaissance de la langue (français) were produced by taking two parts of the Test de français, langue d'enseignement (the first and the fourth), and dividing them so as to produce two "equivalent" test forms, each containing 35 multiple-choice items. Each form of the Test de compréhension en lecture et de connaissance de la langue (français) was divided into two parts. The first part tested reading comprehension by means of two reading passages, each approximately 500 words in length. Students were asked to answer several multiple-choice questions about each passage. These questions involved different types of comprehension: literal understanding of a text; identification of the main idea or objective of an author; and ability to draw conclusions from, and see implications in, what an author says. Although both forms of the test contained two reading passages, Form 1 contained eleven reading comprehension questions, whereas Form 2 contained fourteen.

The second part of the Test de compréhension en lecture et de connaissance de la langue (français) was made up of a variety of exercises testing language achievement. In one type of exercise, the student was given four sentences. His task was to read the sentences and either identify the one containing an error in grammar or indicate that none of the four sentences was in error. Form 1 of the test contained eight items of this type, Form 2 contained seven.

A second language achievement exercise required the student to read a sentence, and, if he decided that it contained an error, to classify that error under one of four headings:

- (a) an error in a noun, adjective or article,
- (b) an error in a pronoun,
- (c) an error in a verb,
- (d) an error in an adverb, preposition or conjunction.

If the student concluded that the sentence was not in error, then he chose the fifth response option for the item. Each form of the test contained five items of this type.

In the final type of language achievement exercise, the student was given an incomplete sentence and asked to choose (from five pairs of words) the pair which best completed the sentence. Form 1 of the test contained eleven of these sentence completion items while Form 2 had only nine.

The Test de composition écrite was a direct translation of the Writing Test developed for administration to Anglophone students. The test offered eight topics, and asked the student to write an expository essay on any one. The student was assured that he could adopt a point of view that differed from or agreed with the point of view expressed in the topic statement, and that his primary concern should be to argue his position logically. A

more detailed description of this test is contained in Appendix A1, which deals with the English language version.

2. TEST APPRAISAL

The tests of competence in the use of French by Francophones were submitted for appraisal to teachers of Grade Twelve and Grade Thirteen courses in the Francophone schools in the study. In all, 34 teachers did the appraisal task. Some responded from the point of view of Grade Thirteen français courses, others from the points of view afforded by Grade Twelve advanced level français courses, Grade Twelve general level français courses, and Grade Twelve basic level français courses. Some individuals responded for courses at more than one level because they were teaching courses at more than one level. As a consequence, the number of appraisals for courses at any one level is smaller than 34--for Grade Twelve basic level courses, the maximum number of responses to an item was only five--although the total number of appraisals across all course levels is larger than 34.

The appraisals must be interpreted in light of the nature of the sample that the 34 teachers comprise. These 34 appraisers were teaching in 13 of the 14 Francophone schools in the study. No test appraisals were done by the teachers in one school. The 14 schools are themselves representative of the population of Francophone schools in the province inasmuch as they constitute a probability sample of the population. This implies that the teachers who appraised the test are representative of the population of français teachers in the province. But a qualification arises because of the way the test appraisals were distributed. Information was received from each school prior to the administration of the tests about the number of persons in the school who taught français courses at the Grade Twelve and Grade Thirteen levels. A corresponding number of test appraisal "kits" was sent to the school. But not all the appraisal forms that were sent were completed and it was not ascertained whether

this was because some appraisers had refused to do the task, as they obviously had in at least one school, or because incorrect information on the number of français teachers had been supplied. For these reasons, it is probably not reasonable to consider the sample of 34 responding teachers as fully representative of the français teachers in the province, although it is relatively so. In any event, generalizing from this sample to a population is hazardous because of the relatively large sampling error that must be associated with such a small number of schools and teachers.

No appraisals were obtained from individuals who were teaching français courses at the first year level in a CAAT college or a university. An appraisal inventory was developed for instructors at this level and it was distributed to a small number of persons but no completed inventories were returned.

The format of the appraisal inventory for the français tests was essentially the same as that for the English tests (see Appendix A1). The first 19 questions of the inventory were directed toward the reading comprehension part of the two forms of the Test de compréhension en lecture et de connaissance de la langue (français). The next 10 questions focussed on the language achievement exercises in the same test. The final 11 questions dealt with the Test de composition écrite. Most questions were not directed toward particular test items. Instead, they asked the appraiser to consider the language skills that were assessed by the tests, the difficulty of different parts of the tests, and the usefulness of the particular testing format employed in the tests. Appraisers were given space to comment in writing at various points throughout the inventory.

The main results of interest as regards the test appraisals are contained in Table A2.1. This table consists of the questions in the inventory, the associated response options and the frequency of response to each option for each level of course.

A detailed discussion of Table A2.1 is not provided. Given the full discussion of results from the appraisal of the tests of English language competence for Anglophones, and given that the picture provided by Table A2.1, in conjunction with the written comments of the appraisers, is similar to that provided by the opinions of the secondary school appraisers of the English tests, it would be redundant to repeat the same points here. A brief summary of the conclusions supported by the appraisals of the français tests seems more appropriate:

- (a) The language objectives examined by the tests--reading comprehension, language achievement and writing--were seen as important.
- (b) Although there was support for the use of a multiple-choice testing format, the consensus was that this kind of test must be supplemented by samples of written work.
- (c) The level of difficulty of the tests was judged to be best suited to students in Grade Thirteen and Grade Twelve advanced level courses, and least appropriate for students in Grade Twelve basic level courses. The tests were judged to be excessively difficult for the latter group.

3. TECHNICAL ISSUES

3.1 Scoring the Test de compréhension en lecture et de connaissance de la langue (français)

As indicated in the instructions to students writing this test, the standard correction for guessing was applied. The multiple-choice items in this test each contained five response options; the standard correction (Lord and Novick, 1968, p.306)

for such items is to assign incorrect answers the weight (-1/4); correct answers are weighted 1 and omitted questions are ignored.

The two forms of this test did not possess equivalent scales, at least in part because of differences in difficulty and reliability. (These differences are described in greater detail in subsequent parts of this section of Appendix A2.) To compensate for these differences, the scores on Form 2 of the test were equated to the scale of scores on Form 1 of the test. This equation was made using the scores of those students who took both forms of the test and the procedure outlined in Appendix D1. Equated scores were used in the studies involving this test, the results of which are described in the main part of this report.

3.2 Scoring the Test de composition écrite

All essays but 50 were scored by three different markers. The 50 essays singled out for special treatment were all scored by all nine markers. The marks on this special set of 50 were used to derive equations for adjusting the scores assigned by different markers. Adjustments were necessary to remove differences among markers in the average mark assigned, the range of marks assigned and the reliability of marking. A single score was obtained for each essay by entering the different marks assigned to it--three or nine--into an equating formula, the specific form of which varied with the number of marks and, in the case of all essays only scored three times, with the particular individuals who marked the essay. The procedure for equating essay marks is described in Appendix D2. All results on these essays that are presented in the main part of this report were derived using the equated scores.

3.3 Difficulty of items in the Test de compréhension en lecture et de connaissance de la langue (français)

The difficulty of each item in each form of the test, as indexed by the average across schools of the percentage of students responding correctly in each Francophone school, is reported in Table A2.2. Percentages are reported separately for Grade Twelve and Grade Thirteen. Given the sampling design of the study, these percentages may be interpreted as estimates for the population of Francophone students defined in this study. (For a definition of this population, see Chapter Two of this report.)

The mean difficulty indices reported in Table A2.2 provide clear support for two conclusions: Form 2 of the test was considerably easier than Form 1, and on the average, Grade Thirteen students performed better than Grade Twelve students. With respect to this latter conclusion, an item-by-item study reveals only two items in each form for which the performance of Grade Twelve students was as good as or better than the performance of Grade Thirteen students. (See the percentages reported in Table A2.2 for items 17 and 18 of Form 1 and items 1 and 2 of Form 2.) (A breakdown of Grade Twelve results, in which students are divided according to their educational plans, is presented and discussed in the main report.) These conclusions are supported by another view of item difficulty, the one presented by the frequency distributions in Table A2.3.

It will be recalled that the Test de compréhension en lecture et de connaissance de la langue (français) was divided into three parts and contained four different types of items. The mean and the standard deviation of the difficulty indices for each item type are reported in Table A2.4. It can be seen from an inspection of the means that the reading comprehension items in the two forms were approximately equivalent in difficulty. Also, the Type 2 language achievement items in Form 1 were approximately equal in difficulty to those in Form 2. (Type 2 items consisted of a sentence; the student was to read it, decide whether or not it contained an error, and classify the error, if

any, under one of four headings.) The main differences in difficulty between the forms are seen to occur with respect to the language achievement items of Types 1 and 3. (Type 1 items consisted of four sentences; students were to read the sentences and select the one, if any, containing a grammatical error. Type 3 items consisted of incomplete sentences that the student was to complete by selecting the most appropriate pair of words.)

There are obvious differences in the difficulty of the different types of items. Regardless of test form and grade level, the reading comprehension items were the easiest of all. Among the language achievement items, those of Type 2 were easiest in Form 1, those of Type 3 were easiest in Form 2.

The main conclusion that can be drawn from this study of the difficulty of items in the Test de compréhension en lecture et de connaissance de la langue (français) is that the two forms differed appreciably in difficulty. Moreover, the two forms, taken together, formed a test that was more difficult than is desirable for the Grade Twelve students as a whole, given that the test was intended to spread students out as much as possible in terms of their performance. (For a test composed of five-option multiple-choice items, an average item difficulty of somewhat higher than 0.60 would be ideal, if the test is to discriminate well among students and scores are to range from near chance level, i.e. 20 per cent correct, to near perfect, i.e. 100 per cent correct.) By this same criterion, the test was on the difficult side even for the Grade Thirteen students.

3.4 Speededness of the Test de compréhension en lecture et de connaissance de la langue (français)

The available information concerning speededness is contained in Tables A2.5 and A2.6. Before this information is considered, however, an aspect of the test administration that bears on speededness needs to be described. Each form of the test was printed in two parts. The first part contained the reading

comprehension passages and related items; the second part contained the three different types of language achievement exercises. The time allowed for the test was 40 minutes, but this time was divided into two 20-minute periods. Students who had not completed the reading comprehension part of the test by the end of the first 20-minute interval were instructed to go on to the language achievement part of the test. Students who finished the reading comprehension part in less than 20 minutes could go on to the language achievement part and students could go back over any part of the test, time permitting, once they had finished the language achievement part. In Form 1, the break between reading comprehension and language achievement occurred between the eleventh and twelfth items. In Form 2, the break occurred between the fourteenth and fifteenth items.

Table A2.5 is a presentation of "not reached" statistics. An item is judged to be not reached if the student fails to respond to it and to all succeeding items in the test. The figures in Table A2.5 suggest that Form 2 was less speeded than Form 1 because, in general, the percentage of students failing to reach a given item in Form 2 is less than the percentage failing to reach the corresponding item in Form 1. The major portion of this difference in speededness seems attributable to the different lengths of the language achievement parts of each test form. Recall that the language achievement part of Form 1 contained 24 items whereas the corresponding part of Form 2 contained only 21 items. Because the instruction to go on to the second part of the test came three items earlier for students working Form 1 than for students working Form 2, it can be argued that the fairest comparison of not reached percentages would involve item 'n', say, of Form 1 and item 'n+3' of Form 2. When this is done, the difference in not reached statistics is relatively small.

A somewhat arbitrary approach to assessing the speededness of a test is to consider both the percentage of students who finish three-fourths of the items and the number of items reached by at least four-fifths of the group. A test is said, by one rule of thumb, to be speeded if less than 100 per cent of the students

reach the three-quarter mark of the test or if less than 80 per cent of the students complete the test. The three-quarter mark of each form (27 items) was reached by 96 per cent or more of the students in each grade. More than 80 per cent of the students completed Form 2, and although a smaller percentage than 80 completed Form 1, if attention is focussed on item 32, the point in Form 1 of fair comparison with the end of Form 2, it will be seen that more than 80 per cent of students reached this item.

A final observation about Table A2.5 concerns the differences between Grade Twelve and Grade Thirteen. In general, the between-grade difference in the percentage of students reaching a given item is small--five percent or less. Moreover, these differences on Form 1 favour neither grade level; for Form 2, the differences lie consistently in the favour of Grade Thirteen.

The evidence in Table A2.5, although it pertains to each test form as a whole, is most indicative of the speededness of the language achievement parts of the test forms. Information that reveals something about the speededness of the reading comprehension parts is contained in Table A2.6. The numbers reported in this table are the percentages of students who failed to respond to a given item, although they did respond to a subsequent item in the test. It can be seen that the number of omissions increases up to the break point between the reading comprehension and language achievement parts of a test form. This increase almost certainly reflects the effect of speededness in the reading comprehension parts of the test forms, although it must be recognized that omissions due to failure to identify a correct answer are confounded in these figures with omissions due to lack of time. Applying the rule of thumb stated earlier, the reading comprehension parts of the test forms are judged to be only very slightly speeded.

Two conclusions can be drawn from the data on speededness that has been presented:

- (a) Both forms of the Test de compréhension en lecture et de connaissance de la langue (français) were somewhat speeded, Form 1 being slightly more so than Form 2. This difference between the forms is, for the most part, explainable by the fact that the break between reading comprehension and language achievement came earlier in Form 1 than Form 2.
- (b) There was very little difference in the degree of speededness of these tests for Grade Twelve students as compared with Grade Thirteen students. What small difference there was tends to indicate that the Grade Thirteen students worked slightly faster than the Grade Twelve students, particularly on Form 2.

3.5 Item discrimination in the Test de compréhension en lecture et de connaissance de la langue (français)

A crude measure of item discrimination is the biserial correlation coefficient between scores on an item (simply 1 for correct and 0 for wrong or omitted) and scores on the total test. In tests designed to spread examinees over the range of possible scores on the test, it is desirable to have relatively high and positive item-total biserial correlation indices, say 0.3 or higher.

The biserial correlation coefficients for each item in both forms of the Test de compréhension en lecture et de connaissance de la langue (français) were computed using the test data from two randomly selected subsamples of the total sample of Francophone students. These groups comprised both SSGD and SSHGD students. The resulting correlations, which we shall call "discrimination indices," are reported in Table A2.7. (Note that decimal points have been omitted from the table.) Only the indices for three items in Form 1 and two in Form 2 fail to exceed the benchmark figure of 0.3. Generally speaking, then, the items in these two test forms may be said to discriminate

adequately among the examinees in the two groups that were studied.

3.6 Distribution and Reliability Statistics for the Test de compréhension en lecture et de connaissance de la langue (français)

Some additional information about the two forms of the test is reported in Table A2.8 and A2.9. It is clear from some of the figures reported in Table A2.8 that both tests were relatively difficult for these students and that Form 1 was more difficult than Form 2. No student achieved the maximum possible mark on either form and the mean scores were exceptionally low.

(The statement about the comparative difficulty of the two forms follows, despite the fact that the results for each form are based on different groups of students, because the groups were equivalent in the sense that they were formed by random assignment of students from the same pool of students.)

The index of reliability and the standard error of measurement are measures of the degree to which test scores might be expected to remain stable over repeated applications of the same or equivalent tests to the same student. The reliability coefficients for these test forms are relatively low; this is due, no doubt, in large part to the heterogeneous nature of the content of the tests. (The correlations among the parts of the test form, as reported in Table A2.9, are only moderately high.) Another reason that the reliability coefficients are low is that the test forms were on the difficult side for Ontario Francophone students; hence scores on the test were not dispersed over the full range of the test score scale as they would be in a test at a more suitable level of difficulty. Despite the relatively low reliability of the test forms and their relatively high standard errors of measurement, the reliability of both instruments is adequate for making the kind of group comparisons and regression analyses demanded in this study.

3.7 Reliability of Scores on the Test de composition écrite

The procedure used to estimate the reliability of essay scoring is described in the section of Appendix A1 entitled "TECHNICAL ISSUES ABOUT THE TESTS". Readers interested in a description of the procedure are referred to Appendix A1.

The figures needed to estimate the reliability of the marking of the français essays are reported in Table A2.10. The remarks that could be made about these results parallel those that are made in Appendix A1 for the English essays.

Application of the procedure described in Appendix A1 leads to an estimate of reliability of 0.77. The corresponding standard error of measurement is 0.72. Both these figures support the conclusion that the scoring of français essays was reliable enough for the purpose to which the scores were put--comparing different groups of students and predicting school grades.

For further information concerning the interpretation that can be given to the coefficient of reliability and the standard error of measurement, the reader is again referred to Appendix A1. The qualifying remarks found there apply with equal validity to the scoring of the Test de composition écrite.

REFERENCES

- Hoyt, C. Test reliability estimated by analysis of variance. Psychometrika, 1941, 6, pp. 153-160.
- Lord, F.M. and Novick, M.R. Statistical theories of mental test scores. Toronto: Addison-Wesley, 1968.

A FREQUENCY TABULATION OF RESPONSES TO THE TEST-INVENTAIRE ESTIMATIF
FOR THE TEST DE COMPREHENSION EN LECTURE ET DE CONNAISSANCE
DE LA LANGUE (FRANCAIS) AND THE TEST DE COMPOSITION ECRITE
(NUMBER OF RESPONDENTS - 34)*

Cette section de l'inventaire a trait à la première partie des deux formules correspondant du test de compréhension en lecture et de connaissance de la langue.

1. Considérez-vous qu'un test de compréhension en lecture --pas nécessairement celui utilisé ici--détermine une composante importante de la connaissance de la langue à ce niveau?

(1) OUI.....

(2) NON

12ème année			
13ème année	niveau avancé	niveau général	niveau de base
18	15	15	4
1			

* See text for explanation of the number of respondents.

TABLE A2.1 (Continued)

12ème année				
	13ème année	niveau avancé	niveau général	niveau de base
Questions 2 à 6: Pour ces questions utilisez les codes-réponses ci-dessous:				
(1) trop facile				
(2) facile				
(3) bien choisi				
(4) difficile				
(5) trop difficile				
2. Si l'on considère les deux formules du test, il y a quatre textes relatifs à la compréhension en lecture. Quelle est votre opinion générale quant au niveau de difficulté de ces quatre textes?	1			
	2	2		
	3	13	8	5
	4	4	6	4
	5		5	3
Déterminez la difficulté de chaque texte séparément en fonction des codes-réponses cités plus haut.				
3. Text sur l'exploration lunaire	1	2		
	2	4	2	2
	3	9	9	6
	4	4	4	4
	5	1	3	1
4. Texte sur la langue française	1			
	2	2		
	3	9	9	5
	4	7	6	5
	5	2	5	5
5. Texte sur la société de consommation	1	1		
	2	2	2	1
	3	13	7	4
	4	4	4	7
	5		2	3

TABLE A2.1 (Continued)

	Response Codes (see previous page)	12ème année			
		13ème année	niveau avancé	niveau général	niveau de base
6. Texte sur le camp de concentration	1				
	2	7	2	2	1
	3	11	7	7	2
	4	2	5	3	2
	5		1	3	
7. Ces quatre textes sont-ils représentatifs de ce que vous attendiez des étudiants de ce niveau quant à la compréhension en lecture?					
(1) OUI.....		15	12	7	3
(2) NON.....		4	2	7	2
Questions 8 à 11: L'un des textes vous semble-t-il ne pas convenir pour des raisons autres que celle de la difficulté?					
8. Texte sur l'exploration lunaire					
(1) OUI.....		4	1	3	1
(2) NON.....		16	14	12	4
9. Texte sur la langue française					
(1) OUI.....		6	2	4	1
(2) NON.....		14	13	11	4

TABLE A2.1 (Continued)

12ème année				
	13ème année	niveau avancé	niveau général	niveau de base
10. Texte sur la société de consommation				
(1) OUI.....	1	1	3	1
(2) NON.....	19	14	12	4
11. Texte sur le camp de concentration				
(1) OUI.....	2	2	5	
(2) NON.....	18	13	10	5
En général, les rubriques sont destinées à tester les compétences de l'étudiant dans les domaines suivants:				
I. compréhension littérale du texte				
II. identification de l'idée principale ou des objectifs de l'auteur				
III. conclusions ou implications				
Les questions 12 à 14 sont relatives au problème de savoir si les étudiants devraient posséder ces trois compétences avant de commencer les cours de français de chaque niveau. Sélectionnez vos réponses à ces questions parmi les options suivantes:				
(1) tous les étudiants				
(2) plus de 75% des étudiants, mais pas tous				
(3) de 51% à 75% des étudiants				
(4) de 26% à 50% des étudiants				
(5) un étudiant ou plus, mais moins de 26%				
(6) aucun				

Response Codes

TABLE A2.1 (Continued)

		12ème année			
		13ème année	niveau avancé	niveau général	niveau de base
12. Combien parmi les étudiants commençant les cours à ce niveau devraient pouvoir comprendre littéralement un texte?	1	11	6	5	1
	2	8	7	6	2
	3	1	2	4	1
	4				1
	5				
	6				
13. Combien d'étudiants devraient pouvoir identifier l'idée principale ou les objectifs de l'auteur?	1	10	5	3	2
	2	10	7	7	
	3		3	5	3
	4				
	5				
	6				
14. Combien d'étudiants devraient pouvoir tirer des conclusions ou dégager des implications?	1	7	3	1	
	2	10	5	8	3
	3	3	4	6	2
	4		3		
	5				
	6				
Les questions 15 à 17 ont trait à une évaluation de l'attention accordée à chacune des compétences en lecture dans les cours que vous enseignez à chaque niveau. Sélectionnez votre réponse à ces questions en choisissant parmi les options suivantes:					
(1) attention très marquée					
(2) attention marquée					
(3) attention modérée					
(4) attention particulière à un étudiant plus faible					
(5) attention nulle					

TABLE A2.1 (Continued)

	Response Codes (see previous page)	12ème année			
		13ème année	niveau avancé	niveau général	niveau de base
15. Quelle attention accordez-vous en lecture à la <u>compréhension littérale</u> du texte?	1	10	7	5	
	2	4	5	8	2
	3	6	3	2	3
	4				
	5				
16. Quelle attention accordez-vous en lecture à l' <u>identification de l'idée principale</u> ou des <u>objectifs de l'auteur</u> ?	1	11	4	3	1
	2	6	7	8	
	3	3	4	4	4
	4				
	5				
17. Quelle attention accordez-vous en lecture au <u>pouvoir de tirer des conclusions</u> au de <u>dégager des implications</u> ?	1	10	2	3	
	2	8	7	6	2
	3	2	6	6	3
	4				
	5				
18. Y a-t-il des compétences importantes en lecture qui n'ont pas été testées dans ces tests mais qui auraient dû l'être?					
(1) OUI.....		6	3	4	1
(2) NON.....		13	12	10	3
19. La formule de questions à choix multiple est-elle une méthode bien fondée pour analyser au moins les trois compétences en lecture mentionnées ci-dessus?					
(1) OUI.....		14	10	10	4
(2) OUI (conditionnel).....		4	1	2	1
(3) NON.....		2	4	3	

TABLE A2.1 (Continued)

	Response Codes (see previous page)	12ème année			
		13ème année	niveau avancé	niveau général	niveau de base
22. Quelle est votre évaluation en ce qui concerne les éléments d' <u>identification d'une phrase fautive?</u>	1 2 3 4	2 16 1 1	3 9 1 2	 11 3 1	 2 3
23. Quelle est votre évaluation en ce qui concerne les éléments d' <u>identification et de classification d'une erreur?</u>	1 2 3 4	 15 3 2	2 9 1 3	 8 6 1	 1 3 1
24. Quelle est votre évaluation en ce qui concerne les éléments de <u>phrases à compléter?</u>	1 2 3 4	 16 2 2	2 11 2	 9 6 	 2 3
25. Quelle est votre évaluation d'ensemble quant au niveau de difficulté de la partie des tests ayant trait à la connaissance de la langue?					
(1) trop facile.....	1		1		
(2) facile.....	2	4	1	1	1
(3) bien choisi.....	3	13	9	7	1
(4) difficile.....	4	3	4	4	3
(5) trop difficile.....	5			3	

TABLE A2.1 (Continued)

<u>CONNAISSANCE DE LA LANGUE</u>		12ème année			
		13ème année	niveau avancé	niveau général	niveau de base
20.	Pensez-vous que la formule de questions à choix multiple soit, en général, satisfaisante pour analyser les compétences des étudiants quant à la langue?				
(1)	OUI.....	8	7	9	3
(2)	OUI (conditionnel).....	6	3	4	1
(3)	NON.....	5	5	2	
21.	Lorsque la formule de questions à choix multiple est utilisée, devrait-elle être complétée par autre chose?				
(1)	OUI.....	12	9	7	1
(2)	NON.....	8	6	8	4
<u>Questions 22 à 24:</u> En termes de convenance et de difficulté en tant que mesures de compétence linguistique, comment évalueriez-vous les éléments de: "identification d'une phrase fautive", "identification et classification d'une erreur" et "phrases à compléter" qui apparaissent dans ces tests? Sélectionnez votre réponse à ces questions en choisissant parmi les options suivantes:					
(1)	appropriés mais trop faciles				
(2)	tout à fait appropriés				
(3)	appropriés mais trop difficiles				
(4)	inappropriés pour des raisons autres que la difficulté				

Response Codes

TABLE A2.1 (Continued)

		12ème année			
		13ème année	niveau avancé	niveau général	niveau de base
26.	Quelle est votre évaluation des éléments des tests relatifs à la connaissance de la langue du point de vue de l'accent mis sur l'usage, le style, la grammaire, la structure des phrases, les expressions idiomatiques?				
(1)	ils fournissent un bon équilibre en testant des secteurs importants de la connaissance de la langue.....	16	11	13	4
(2)	bien que testant un certain nombre de secteurs importants, l'accent n'est pas mis de façon équilibrée sur les différents éléments.....	1	3		
(3)	des secteurs importants de la connaissance de la langue sont omis ou testés trop superficiellement.....	2	1	1	1
27.	Combien d'étudiants devraient avoir les compétences estimées dans les tests de connaissance de la langue en <u>entrant</u> dans un cours de niveau donné?				
(1)	tous.....	6	2	2	1
(2)	plus de 75%, mais pas tous.....	9	8	3	
(3)	de 51% à 75%.....	5	4	7	2
(4)	de 26% à 50%.....		1	3	2
(5)	un ou plus, mais moins de 26%.....				
(6)	aucun.....				

TABLE A2.1 (Continued)

		12ème année			
		13ème année	niveau avancé	niveau général	niveau de base
28.	Quelle attention accordez-vous au développement des compétences requises pour les tests de connaissance de la langue?				
(1)	attention nulle--les compétences requises sont sans intérêt.....				
(2)	attention très marquée.....	4	7	3	1
(3)	attention marquée.....	9	5	7	2
(4)	attention modérée.....	5	3	3	2
(5)	attention particulière à un étudiant plus faible.....	2		1	
(6)	attention nulle--les compétences requises sont trop avancées.....			1	
29.	Combien d'étudiants qui terminent avec succès les cours de français à ce niveau devraient avoir les compétences évaluées par les tests de connaissance de la langue?				
(1)	tous.....	10	7	4	1
(2)	plus de 75%, mais pas tous.....	7	6	5	2
(3)	de 51% à 75%.....	2	2	3	1
(4)	de 26% à 50%.....			3	1
(5)	un ou plus, mais moins de 26%.....				
(6)	aucun.....				

TABLE A2.1 (Continued)

TEST DE COMPOSITION ECRITE		12ème année			
		13ème année	niveau avancé	niveau général	niveau de base
NOTE: Dans vos réponses aux questions 30 à 35, vous ne vous référerez pas en particulier à l'essai proposé.					
30.	Quelle importance accordez-vous à un exemple de la composition de l'étudiant dans une évaluation de la compétence au niveau de la langue?				
(1)	importance très marquée...	8	8	2	1
(2)	importance marquée.....	10	5	8	1
(3)	importance modérée.....	1	1	4	3
(4)	importance nulle.....			1	
31.	Comment considérez-vous l'utilisation, à la fois d'un test de connaissance de la langue avec questions à choix multiple et un exemple de composition, dans l'évaluation de la compétence au niveau de la langue?				
(1)	la formule de questions à choix multiple est satisfaisante en elle-même.....			1	
(2)	le recours aux deux formules est important....	15	12	14	4
(3)	la formule d'un exemple de composition est satisfaisante en elle-même.....	4	1		
(4)	aucune des deux formules n'est satisfaisante.....		1		1

TABLE A2.1 (Continued)

	12ème année			
	13ème année	niveau avancé	niveau général	niveau de base
33. Combien d'étudiants devraient pouvoir écrire un essai correct - du type présenté ici - à son entrée dans les cours de français au niveau donné?				
(1) tous.....	11	2		
(2) plus de 75%, mais pas tous.....	7	8	7	2
(3) de 51% à 75%.....	2	3	5	1
(4) de 26% à 50%.....		1	2	
(5) un ou plus, mais moins de 26%.....		1		2
(6) aucun.....			1	
34. Quelle attention accordez-vous, dans votre enseignement, au développement de la compétence des étudiants dans ce type de composition écrite?				
(1) attention très marquée....	7	4	1	
(2) attention marquée.....	7	8	4	2
(3) attention modérée.....	6	3	9	3
(4) attention nulle.....			1	
35. Combien d'étudiants qui terminent avec succès les cours de français à ce niveau devraient pouvoir écrire un essai correct du type de ceux proposés dans ce test?				
(1) tous.....	12	4	1	
(2) plus de 75%, mais pas tous.....	6	10	6	2
(3) de 51% à 75%.....	2	1	5	1
(4) de 26% à 50%.....			1	2
(5) un ou plus, mais moins de 26%.....			1	
(6) aucun.....			1	

TABLE A2.1 (Continued)

		12ème année			
		13ème année	niveau avancé	niveau général	niveau de base
36.	Y a-t-il d'autres modes de composition écrite qui, du point de vue d'une culture générale, seraient aussi ou plus importants que celui utilisé ici?				
(1)	OUI.....	6	9	11	3
(2)	NON.....	12	5	4	2
NOTE: Répondez aux questions 37 à 40 en tenant compte de la composition écrite proposée.					
37.	La composition écrite était-elle située à un niveau raisonnable de difficulté pour des étudiants suivant les cours de ce niveau?				
(1)	OUI.....	19	15	8	2
(2)	NON.....	1		7	3
38.	Le fait de n'avoir proposé qu'un seul type de composition écrite était-il de nature à défavoriser les étudiants?				
(1)	OUI.....	4	4	5	1
(2)	NON.....	16	11	10	4
39.	Compte-tenu de fait de n'avoir proposé qu'un seul type de composition écrite, que pensez-vous de la variété des sujets?				
(1)	bonne.....	13	11	11	4
(2)	satisfaisante.....	7	4	2	1
(3)	non satisfaisante.....			2	

TABLE A2.1 (Continued)

		12ème année			
		13ème année	niveau avancé	niveau général	niveau de base
40.	Numérotez les critères suivants de 1 à 5, par ordre d'importance décroissante, pour l'évaluation de ce type de composition écrite. Inscrivez les chiffres dans les cinq espaces prévus à cet effet dans chacune des 4 colonnes.				
Structure générale	1	6	9	4	1
	2	2	1	3	
	3	3	2	1	1
	4	7	2	5	3
	5	3	2	1	1
Présentation logique des arguments	1	5	3	5	
	2	5	6	4	2
	3	1	2	1	1
	4	8	4	3	1
	5	2	1	1	2
Style, principalement dans la phrase	1	3	1		
	2	7	7	3	2
	3	9	3	6	3
	4	2	4	5	1
	5		1		
Technique de la langue: grammaire, usage, mécanique	1	4	3	4	4
	2	5		1	1
	3	4	6	5	
	4	4	5	2	
	5	4	2	2	1
Choix des mots	1	2	2		1
	2	1	1	2	1
	3	4	2	1	1
	4	3	2	1	1
	5	11	9	10	2

TABLE A2.2

Percentages of Correct Answers to Items in the Test de
compréhension en lecture et de connaissance de la
langue (français) by Form of Test and Grade Level

Item No	Form 1		Form 2		Item No	Form 1		Form 2	
	SSGD	SSHGD	SSGD	SSHGD		SSGD	SSHGD	SSGD	SSHGD
1	57	72	76	76	21	25	42	48	51
2	51	67	65	78	22	61	70	63	69
3	57	68	75	83	23	19	20	35	35
4	51	70	57	67	24	36	42	29	33
5	73	74	60	83	25	45	67	37	45
6	53	75	57	69	26	44	60	12	17
7	50	67	82	95	27	10	19	26	48
8	16	26	24	26	28	36	34	50	66
9	57	58	58	68	29	7	17	63	74
10	45	52	33	51	30	19	32	46	57
11	25	38	34	42	31	12	21	57	76
12	31	44	20	28	32	25	30	26	48
13	25 ^a	33 ^a	25	36	33	47	50	34	54
14	-	-	19	30	34	16	30	26	41
15	54	64	38	51	35	14	20	31	53
16	14	21	28	29	-----				
17	36	52	20	26	Mean	34	44	42	53
18	14	9	36	44	S.D.	18	20	18	19
19	19	15	32 ^a	43 ^a					
20	18	26	-	-					

^aNo estimate of difficulty is reported for items judged to have two correct answers.

TABLE A2.3
Frequency Tabulations of Proportions Reported
in Table A2.2

<u>Percentage</u>	<u>Form 1</u>		<u>Form 2</u>	
	<u>SSGD</u>	<u>SSHGD</u>	<u>SSGD</u>	<u>SSHGD</u>
90-99	-	-	-	1
80-89	-	-	1	2
70-79	1	5	2	4
60-69	1	6	4	5
50-59	8	4	5	6
40-49	4	3	2	7
30-39	4	6	8	4
20-29	4	6	10	4
10-19	11	3	2	1
00-09	1	1	-	-
n	34	34	34	34

TABLE A2.4

Means and Standard Deviations of Percentages of Correct
Answers for Each Different Type of Item in the Test
de compréhension en lecture et de connaissance de
la langue (français)

Item Type	Form 1				Form 2						
	No. of Items	SSGD		SSHGD		No. of Items	SSGD		SSHGD		
		Mean	S.D.	Mean	S.D.		Mean	S.D.	Mean	S.D.	
Reading Comprehension	11	48	15	61	15	.	14	49	22	59	23
Language Achievement											
Type 1 ^a	7	28	13	34	19	.	6	34	9	41	10
Type 2	5	32	16	40	17	.	5	35	16	40	17
Type 3	11	25	15	35	16	.	9	40	14	57	11

^aThe number of items of this type in each form of the test was one more than the number reported here but one item in each form was excluded from this analysis because it was double keyed.

TABLE A2.5

Percentage of Students at Each of Two Grade Levels
Not Reaching the Items Appearing Towards the End
of Each Form of the Test de compréhension en
lecture et de connaissance de la langue
(français)

Item No	Form 1		Form 2	
	SSGD	SSHGD	SSGD	SSHGD
20	0	0	- ^a	- ^a
21	1	0	0	0
22	1	0	1	0
23	1	0	1	0
24	1	0	1	0
25	2	1	2	0
26	3	2	2	0
27	3	4	3	1
28	5	5	3	1
29	6	9	4	2
30	8	11	5	2
31	13	13	7	3
32	17	18	7	4
33	22	20	10	7
34	24	25	13	9
35	32	34	18	13

^aThis item of Form 2 was judged to be double-keyed and consequently was omitted from the item analysis.

TABLE A2.6

Percentage of Students at Each of Two Grade Levels Who
Omitted the Items in the Two Forms of the Test de
compréhension en lecture et de connaissance
de la langue (français)

Item No	Form 1		Form 2		Item No	Form 1		Form 2	
	SSGD	SSHGD	SSGD	SSHGD		SSGD	SSHGD	SSGD	SSHGD
1	3	2	0	1	19	4	7	3 _b	7 _b
2	1	1	0	0	20	11	9	-	-
3	3	5	1	1	21	14	13	2	8
4	4	1	1	1	22	11	8	5	8
5	2	2	0	0	23	15	17	9	10
6	5	3	2	1	24	16	14	11	12
7	7	4	2	0	25	1	1	9	10
8	8	12	3	2	26	3	7	13	13
9	14	16	6	3	27	1	1	3	1
10	14	13	9	4	28	7	12	1	2
11 _a	19	17	11	6	29	2	4	1	3
12	5	4	13	9	30	1	3	1	4

TABLE A2.6 (continued)

Item No	Form 1		Form 2		Item No	Form 1		Form 2	
	SSGD	SSHGD	SSGD	SSHGD		SSGD	SSHGD	SSGD	SSHGD
13	5 ^b	7 ^b	13	6	31	5	6	3	1
14	-	- ^a	15	11	32	1	4	2	6
15	5	7	2	5	33	1	1	5	6
16	4	8	4	7	34	1 ^c	1 ^c	1 ^c	2 ^c
17	5	4	3	5	35	0	0	0	0
18	3	6	3	8					

^aThis line marks the break between the items in the reading comprehension part of the form and those in the language achievement part.

^bResults for this item are not reported as it is judged to be double keyed.

^cAll omissions to the last question in the test are arbitrarily classed as Not Reached.

TABLE A2.7

Indices of Discrimination for Items in Both Forms of
the Test de compréhension en lecture et de
connaissance de la langue (français)

<u>Item No</u>	<u>Form 1</u> <u>(N=248)^a</u>	<u>Form 2</u> <u>(N=249)^b</u>
1	54	45
2	47	62
3	51	53
4	43	54
5	31	75
6	50	35
7	45	48
8	42	20
9	32	37
10	53	62
11	51	39
12	40	52
13	39	30
14	- ^c	49
15	44	49
16	57	45
17	55	53
18	27	46
19	21	35
20	45	- ^c
21	60	48
22	37	48
23	38	32
24	37	16
25	70	53
26	52	41
27	37	58
28	15	55
29	64	31
30	51	42

TABLE A2.7 (continued)

<u>Item No</u>	<u>Form 1</u> <u>(N=248)^a</u>	<u>Form 2</u> <u>(N=249)^b</u>
31	52	80
32	40	73
33	36	63
34	55	49
35	60	66
Mean	45	48
S.D.	12	14

Note: Decimal points have been omitted from indices.

^aThe students in this group wrote both forms of the test in the order Form 1--Form 2; only the results for Form 1 are reported here.

^bThe students in this group wrote both forms of the test in the order Form 2--Form 1; only the results for Form 2 are reported here.

^cThese items were judged to be double-keyed and were therefore omitted from subsequent analysis.

TABLE A2.8

Some Statistics Describing Both Forms of the
Test de compréhension en lecture et de
connaissance de la langue (français)

	<u>Form 1</u>	<u>Form 2</u>
Number of Examinees	248 ^a	249 ^b
Number of Test Items	34	34
Mean	9.23	12.09
Standard Deviation	6.21	7.10
Highest Score	29.00	29.00
Lowest Score	-3.25 ^c	-4.75 ^c
Reliability	0.76	0.80
Standard Error of Measurement ^d	3.00	3.11

^aThese examinees took both forms of the test in the order Form 1--Form 2.

^bThese examinees took both forms of the test in the order Form 2--Form 1.

^cNegative scores arise as a consequence of the application of a correction-for-guessing; in this case, each wrong answer was scored (-1/4) and each correct answer was scored (+1).

^dComputed by applying a formula due to Hoyt (1941).

TABLE A2.9

Correlations Among Scores on the Subtests and Total
Test of Both Forms of the Test de compréhension en
lecture et de connaissance de la langue (français)

<u>Test Part</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
1. Reading Comprehension Subtest	-	41	55	84
2. First Language Achievement Subtest	38	-	49	76
3. Second Language Achievement Subtest	48	52	-	83
4. Total Test	80	79	82	-

Note: Decimal points have been omitted. Correlations below the diagonal are for Form 1, those above the diagonal are for Form 2. For correlations below the diagonal, the sample size was 248; for correlations above the diagonal, the sample size was 249.

TABLE A2.10

Summary of the Analysis of Variance of Scores Assigned
to 50 of the Tests de composition écrite by Five Markers on
Each of Two Occasions

Source of Variance	Degrees of Freedom	Mean Square	Estimate of Variance Component
Essays	49	19.24	1.74
Markers	4	25.95	0.21 ^a
Occasions	1	0.66	-
Essays x Markers	196	1.67	0.49
Essays x Occasions	49	0.86	0.03
Markers x Occasions	4 ^b	3.90	0.06
Residual	190 ^b	0.70	0.70

^aThis was a negative number, subsequently treated as zero.

^bFor 50 essays, 5 markers and 2 occasions, this number should be 196, making a total of 499 d.f. Unfortunately six scores in the total set of 500 were missing, either because of administrative error--occasionally an essay was not included in the bundle sent to a marker--or because of marker error--occasionally an essay was returned unscored. In the analysis of variance, each missing score was estimated as the sum of the grand mean and the main effects for the essay, the marker and the occasion for which the score was missing.

APPENDIX A3

TECHNICAL REPORT ON THE TESTS OF FRENCH AS A SECOND LANGUAGE

This set of four tests was designed to see how well students can read, understand, write and speak French. The tests were originally prepared for use in the IEA world-wide survey of achievement in French (Carroll, 1975) and were adapted slightly for use in Project II. The tests were administered to a selected group of those students taking one or more Grade Thirteen courses in French. All students chosen for testing were assigned the Reading Test and all but at most four students per school were assigned the Listening Test. Those students in each school who were not assigned the Listening Test were assigned either the Writing Test or the Speaking Test, but not both.

This report consists of three main sections: a description of the content of each test; a summary of the test appraisals done by a group of secondary school teachers of Grade Thirteen French courses and a group of university instructors of first year French courses; and a presentation of information about some technical matters related to the tests.

1. DESCRIPTION OF TEST CONTENT

1.1 Reading Test

This was a test of ability to understand written French. Students were allowed 30 minutes to complete a total of 39 multiple-choice questions. The test was divided into two parts.

The first part consisted of 17 incomplete sentences. Students were to choose the answer that best completed each sentence from among the four choices given. In the second part of the test, there were six reading passages ranging in length from approximately 60 words to approximately 150 words. Each passage was accompanied by three or four or five questions or incomplete statements. On the basis of the material contained in the passage, students were to choose the best answer from among the four given for each item.

1.2 Listening Test

This was a test of ability to understand spoken French. It took about 25 minutes. A total of 34 multiple-choice items were presented on tape. Answers to each question were chosen from the four possibilities printed in the test book. The Listening Test had five parts. In the first part, students heard a series of statements. Each statement described one of four pictures labelled A through D in the test book. Students were to choose the picture that best fit the statement. Part Two consisted of a series of remarks or questions spoken by the tape-recorded voice. After each remark or question, students selected the most appropriate response from among the four offered in the test book. In the third part of the test, students listened to a series of short conversations. (One person asked a question or made a statement, and another replied.) After each conversation, students were to select, from among the four statements printed in their test books, the one that was correct according to the conversation. The fourth part of the Listening Test consisted of a series of short broadcasts or announcements, each followed by a question--still on tape--about what had been said. The best answer to the question was to be chosen from among four choices printed in the test book. Finally, in Part Five of the test, students heard relatively long dramatic scenes or conversations. Each scene or conversation was repeated and then several questions were asked about it. From among the four choices given in the

test book, students were to select the best answer to each question.

The number of items in each of Parts One to Five of the Listening Test was 7, 9, 8, 5 and 5 respectively.

1.3 Writing Test

This test of ability to write French was divided into two separately-timed parts. Fifteen minutes were allowed for the first part and ten minutes for the second. Part One of the test contained two types of item. The first type presented a sentence in which one word had been replaced by a blank space. Students were to complete the sentence by writing in the blank space a single French word, correct in both form and meaning. There were 26 items of this type. The second type of item consisted of a pair of sentences. The first sentence had two or three words underlined; the second sentence had a corresponding number of blank spaces. Students were to use the underlined words in the first sentence, making whatever changes in form were required, to fill the blank spaces in the second sentence. In the test there were six items of this second type. In Part Two of the Writing Test, students were asked to write a short composition, approximately one-half page in length, on a given theme. Various aspects of the theme were listed, and students were asked to cover all of them in the order in which they were listed, and to omit none.

1.4 Speaking Test

This test of ability to speak French contained four parts and required about 25 minutes to administer to an individual student. All responses were spoken into a microphone and recorded on tape. In the first part of the test the student heard a series of French sentences spoken twice. After the second presentation of each sentence, the student was to repeat it. The second part of

the test consisted of a set of pictures and a question about each. The student was to reply to each question in a complete sentence, and not in a single word. The third part of the test presented the student with a prose passage of approximately 150 words. He was given three minutes to study the passage and then asked to read it aloud. The last part of the test consisted of two exercises. In the first exercise, the student chose a sequence of three pictures from among three such sequences. Once his choice had been made, he was given one minute to prepare, and then 30 seconds to present, an oral description of the events in the picture sequence. In the second exercise, the student was asked to choose one of three pictures and then to describe the following: the probable action that took place before the events depicted, the situation actually depicted, and the probable outcome of the situation. Again, the student had one minute for preparation and 30 seconds for oral presentation.

2. TEST APPRAISAL

The Tests of French as a Second Language were appraised by 64 secondary school teachers and 18 university instructors. To qualify as an appraiser, a secondary school teacher had to be teaching a Grade Thirteen French course in one of the 53 Anglophone schools in the study. Teachers from only 46 of the schools did the appraisals. University instructors qualified as appraisers if they taught courses in French to first year students in one of the universities chosen for inclusion in Secondary-Postsecondary Interface Project III--The Nature of Programs. Because no effort was made to draw a probability sample of university instructors of French and because it is not known how many secondary school teachers refused to appraise the tests, it is not possible to use the opinions of the individuals who did serve as appraisers as a basis for inferring what the population of instructors across the province would say about the tests. These results are presented only as the collective opinions of a

group of 82 individuals qualified by virtue of their professions to serve as appraisers of these tests.

2.1 Reading Test

As their first task, appraisers were asked for judgments on each of the 17 items in Part A of the Reading Test. Secondary school appraisers were asked to judge whether the reading knowledge of French required to answer each of these multiple-choice questions was "old knowledge" that students should have on entry to a Grade Thirteen course, "new knowledge" that should be learned by all or some specified fraction of students taking a Grade Thirteen course, or "new knowledge" that no student in a Grade Thirteen course could be expected to acquire. University instructors were asked to make similar judgments with respect to students in first year university French courses. A tabulation of the responses given by secondary and postsecondary appraisers is contained in Table A3.1.

It is clear from Table A3.1 that for most items in Part A of the Reading Test there was a measure of disagreement amongst the appraisers as to whether or not an item measured knowledge that all or most students could be expected to have at the end of secondary school, just prior to entry into university. The only items which at least 75 per cent of the secondary school appraisers judged that 75 per cent or more students should have the knowledge to answer correctly were items 1 through 7, inclusive, and items 12 and 13. Of these, only items 1 through 5, inclusive, and item 12 were judged by at least 75 per cent of postsecondary appraisers to tap knowledge that students should have on entry to university courses. Viewing Part A of the Reading Test in this way leads to the conclusion that only one-half or fewer of the items in this part of the test measured knowledge that most appraisers felt most students should have at the interface between secondary and postsecondary education.

The low opinion that the appraisers had of Part A is further confirmed by the tabulation of responses to a further question: "Do you consider that Part A of the Reading Test, taken as a whole, assesses an important component of French language achievement at the Grade Thirteen level?" Of the 61 secondary school appraisers who responded to this question, only 33 responded yes. Thirteen of the 18 university appraisers gave an affirmative answer. Among the reasons offered for responding negatively to this question were the following: some items tested rarely used forms which need not be stressed, some items were too easy, the range of material covered was too limited, and some items were too "literary".

In the appraisal of Part B of the Reading Test, respondents were asked to assess the difficulty of the six reading passages contained in this part of the test. Results based on this assessment are given in Table A3.2. The secondary school appraisers judged passages 1, 2 and 4 to be on the easy side and passages 5 and 6 to be on the difficult side. Generally speaking, this assessment was confirmed by the university appraisers. In the assessment of appropriateness, discounting difficulty, only the fifth passage was labelled inappropriate by a substantial number of appraisers. This passage was described by several of the appraisers who offered written comments as uninteresting and confusing, even to those who know and use French as their first language.

Appraisers were then asked about the kinds of knowledge tapped by the multiple-choice questions based on the reading passages. Roughly speaking, the questions were designed to assess:

- (a) the ability to understand the literal meaning of a passage,
- (b) the ability to identify the main idea or purpose of a passage, and

- (c) the ability to draw inferences or implications from what is written in a passage.

Respondents estimated the proportion of students who should have these capabilities, both on entry to, and on successful completion of, courses at the Grade Thirteen and first year university levels. A tabulation of responses to these questions is contained in Table A3.3. The results that bear on the interface are the estimates from secondary school appraisers for students who have successfully completed Grade Thirteen courses and those from university appraisers for "students entering first year university courses. Fewer than 75 per cent of the secondary appraisers expected 75 per cent or more of students at the interface--that is, students who had successfully completed a Grade Thirteen course in French--to be able to draw inferences from, and see implications in, the ideas they encountered in French passages like those in the Reading Test. But 75 per cent or more of these appraisers did expect 75 per cent or more of students at the interface to be able to read for literal understanding and to possess the ability to identify the main purpose or idea of a passage. The opinion university appraisers held of the abilities of students at the interface was somewhat lower. The only skill that at least 75 per cent of the university appraisers indicated was possessed by 75 per cent or more of students entering first year university courses in French was reading for literal understanding.

Appraisers were asked to rate the degree of emphasis they gave to the aforementioned three reading skills in their Grade Thirteen and first year university French courses. The responses are tabulated in Table A3.4. Regardless of level, a large majority of appraisers indicated that they gave all three skills some degree of emphasis. This fact supports the inclusion in the Reading Test of items that attempted to assess these skills.

The final three questions in the Reading Test appraisal concerned the reading skills not assessed by the test, the usefulness of the multiple-choice format for assessing the skills

tapped by the Reading Test, and whether or not the kind of test represented by the Reading Test was used by appraisers in their Grade Thirteen and first year university courses. A tabulation of responses to these questions is given in Table A3.5. Although the majority of secondary school and university appraisers indicated that there were no important reading skills that were not assessed by the Reading Test, approximately one-fourth of the appraisers did feel there were important omissions. Two skills these individuals felt were not being assessed were the ability to make grammatical distinctions and the ability to detect differences in style and tone of writing. Even the abilities to identify main ideas and draw inferences were mentioned because several appraisers felt these skills were not well assessed by the test despite the fact that it was intended to measure them.

On the question of whether or not the multiple-choice format was a reasonable way to assess at least those reading skills tapped by the Reading Test, only four appraisers expressed a negative view. Of the vast majority who said YES, in response to this question, somewhat more than half qualified their answer. In their written comments, a number of appraisers said that multiple-choice questions should be supplemented by questions demanding an answer composed by the student. Other commentators mentioned the need for supplementary oral testing to clarify a student's multiple-choice responses. And several appraisers raised the issue of the possible influence of guessing on multiple-choice test scores.

The last question for which a tabulation appears in Table A3.5 concerned the use by the appraisers of tests similar to the Reading Test. The number of secondary school appraisers who reported using this kind of test, at least occasionally, was about twice the number who reported not using it at all. The majority of university appraisers indicated that they did not use tests of this nature.

The picture provided by the appraisals of the Reading Test can be outlined roughly as follows:

- (a) Part A of the test, consisting of incomplete sentences and four alternative completions, was viewed quite negatively. Only one-half or fewer of the items in this section were judged by most appraisers to be suitable for assessing some aspect of the reading competence of students at the interface.
- (b) Part B of the test, consisting of reading passages and related questions, was viewed much more favorably. The reading abilities assessed in this part of the test were ones that students were judged to possess in varying degrees at the interface and that instructors said they emphasized to a greater or lesser extent in their courses. The reading passages themselves were generally seen as satisfactory as far as difficulty was concerned, although various complaints were lodged against one passage or another.
- (c) The multiple-choice question format employed in the Reading Test was regarded as acceptable by all but a few of the appraisers.

2.2 Listening Test

Appraisers were asked whether students should have acquired the knowledge of French needed to answer each item in the Listening Test either before entering or during a Grade Thirteen or first year university course. Knowledge students would have on entry to a course was to be classified as "old"; knowledge that would be learned during a course was to be classified as "new". For each item testing new knowledge, secondary school appraisers were asked to estimate the percentage of students who would master the new material during the Grade Thirteen year. A tabulation of responses is given in Table A3.6. Those responses reflecting a

judgment that 75 per cent or more of students at the interface should have the knowledge required to answer an item lie to the left of the dotted lines in Table A3.6; one line for responses in the secondary school appraisals, the other for responses in the university appraisals. For 26 of the 34 items, 75 per cent or more of the secondary school appraisers indicated that 75 per cent or more of the students at the interface should have the knowledge required to answer correctly. Applying the same criteria to the responses of university appraisers leads to the identification of 23 items (all but one of which are in the set of 26 items identified in the analysis of responses from secondary school appraisers) testing knowledge which should be "old" by the time a first year university French course is begun.

It will be recalled from the description of the Listening Test given in the first part of this technical report that the test was divided into five parts. The line separations in Table A3.6 divide the items as they were divided among the five parts of the test. Part Four is the only part of the test for which a majority of the items (3 of 5) were judged by less than 75 per cent of both the secondary school appraisers and the university appraisers to require knowledge most students at the interface should have.

Additional reactions to the Listening Test were obtained through a series of four relatively general questions. These questions and tabulations of responses to them are given in Table A3.7. It is apparent from the results for the first of these questions that a majority of the secondary school instructors who appraised the tests and all but one of the university instructors placed heavy or moderately heavy emphasis on the development of listening skills in their French courses. Responses to the second general question indicate that a substantial majority of both the secondary school and university instructors could think of no important listening skills other than those included in the test. Those respondents who did think that important skills were omitted went on to mention, in written comments, such things as the ability to adjust one's listening to the speed of a native

speaker, the ability to discriminate various vowels and consonants, and the ability to identify the sound of a word in a group of similar-sounding words.

In response to the question about the use of a tape-recorded script and multiple-choice questions in testing for achievement of listening skills, all but one respondent indicated that this kind of test was acceptable, although a number of respondents qualified their response. Several appraisers said that use of a "mechanical" recorded voice creates an artificial testing situation compared with use of a "live" reader. None of these comments showed an awareness of the need in survey testing to provide the kind of standardization of administrative conditions that use of a tape recorder provides but use of "live" readers does not.

On the question about using tests similar to the Listening Test in their own courses, the majority of respondents indicated that they made some use of such tests. Those who said they made no use of this kind of test indicated that they did not have access to suitable test tapes or that they did not approve of multiple-choice questions.

The results of the appraisal of the Listening Test suggest the following conclusions:

- (a) Approximately three-fourths of the items in the test were ones that most students at the interface should have been able to answer correctly.
- (b) The listening skills assessed by the test are important; most of the appraisers strove to teach these skills to some degree in their own courses. Moreover, most instructors were of the opinion that no important listening skills were neglected in the test.

- (c) Reasonable procedures were employed in the test--tape-recorded voice and multiple-choice questions--to achieve standardized testing conditions.

2.3 Writing Test

This test was composed of two separately-timed sections. Questions in the appraisal inventories dealt with each section of the test separately and the results are reported separately here.

Part One of the Writing Test consisted of 32 sentence completion items (see Section 1.3 of this report). Appraisers were asked to judge each item in terms of whether or not the knowledge required to complete the statement was knowledge that students in Grade Thirteen or first year university courses could be expected to have, and if so, whether they would have acquired the knowledge before entering the course, in which case it would be "old" knowledge, or while in the course, in which case it would be "new" knowledge. Tabulations of the judgments of the two groups of appraisers are given in Table A3.8. Note that a dotted line in the tabulation for each group of respondents separates those responses reflecting a judgment that 75 per cent or more of students at the interface should know enough French to answer a test question correctly from responses reflecting a judgment that fewer than 75 per cent of students at the interface should have the requisite knowledge to answer a question correctly.

The results in Table A3.8 clearly indicate that the large majority of both secondary school and university level appraisers thought that students at the interface should know enough French to answer the questions in Part One of the Writing Test. Only four of the 32 items received the endorsement of fewer than 13 university appraisers as items that students entering first year courses should be able to answer. (The number 13 was used in this analysis because it represents approximately 75 per cent of 18, the number of university appraisers.) All 32 items were

judged by at least 75 per cent of the secondary school appraisers to tap knowledge that 75 per cent or more of students at the interface should have. In view of these figures, it is tempting to conclude that the questions in Part One of the Writing Test were highly appropriate for students at the interface. What is disturbing, however, is the fact that most of the questions tapped what most secondary school appraisers classified as "old" knowledge. There were only a few items that even as many as 25 per cent of the secondary school appraisers judged to be testing knowledge that would be acquired in Grade Thirteen. From this point of view, Part One of the Writing Test could be criticized for focussing on inappropriately elementary knowledge.

For results on the appraisal of Part Two of the Writing Test--that part in which students were asked to write a composition--reference is made to Tables A3.9, A3.10 and A3.11. Table A3.9 is based on responses to the request for an estimate of how many students entering and successfully completing French courses in Grade Thirteen and first year university should be able to write an acceptable composition of the type found in the Writing Test. More than two-thirds of the secondary school and university appraisers were of the opinion that 75 per cent or more of students at the interface should be able to write an acceptable composition of this type.

In response to a question about the degree of emphasis given to the development of student competence in writing French compositions (see Table A3.10), the large majority of Grade Thirteen and university instructors reported the expenditure of some effort in this direction, and more than half the individuals in each group said they gave this objective either heavy or moderately heavy emphasis. The ability to write compositions in French is clearly an important skill in the opinion of these appraisers.

One question in the appraisal asked whether or not there were types of writing other than the kind assessed in the Writing Test that students in Grade Thirteen and first year university

courses should be able to do (see Table A3.11, Question 1). Most appraisers said that there were. The following additional kinds of writing were suggested: translations, formal essays in which literary evaluations are made, resums, stories, character sketches, answers to questions about textual material read by the students, creative essays, letters.

Appraisers were then asked for an assessment of the difficulty of the composition assigned in Part Two of the Writing Test. Most of them agreed that it was at a reasonable level of difficulty (see Table A3.11, Question 2), although there were several comments to the effect that the assignment was too easy or too juvenile.

On the matter of the criteria used to judge the compositions that students wrote (see Table A3.11, Question 3), the majority of appraisers felt that there were other criteria of importance in addition to total number of clauses, grammar, extent of vocabulary and accuracy in the use of vocabulary (excluding spelling and use of accents). The other criteria that appraisers suggested included style, clarity, maturity of ideas, originality, creativity, organization and logic, coherence, ability to interest the reader, flair, spelling and accents, variety of sentence structure, ease of expression.

The appraisals of Part Two of the Writing Test suggest the following conclusions:

- (a) The ability to write a composition in French is important. Moreover, it is a skill that most of the appraisers tried to foster in their students.
- (b) The writing task that was assigned was at a reasonable level of difficulty, if slightly on the easy side, for students at the interface.

- (c) There are types of writing other than the one included in the test that students should be able to do and there are important criteria for judging compositions in addition to the ones used with the test.

Appraisers gave a summary evaluation of the Writing Test (see Table A3.11, Question 4) in terms of whether or not they used similar instruments in their own work. Most agreed that they did, at least occasionally.

2.4 Speaking Test

This instrument was divided into four parts. In the first part, students were asked to repeat a series of French phrases that were spoken to them by a tape-recorded voice. A tabulation of the responses of appraisers to the items in Part One is given in Table A3.12. It is clear from this tabulation that most students at the interface could be expected to have the knowledge needed to succeed on this part of the test. In fact, most of the secondary school appraisers judged the content of this part of the test to be focussed on material learned prior to Grade Thirteen.

A summary of the assessment of Part Two of the test is given in Table A3.13. Items in this part of the test consisted of cartoons and the student's task was to answer in French a question asked in French about each cartoon. It is apparent from Table A3.13 that the items in this part of the test focussed on knowledge that students at the interface should have. In fact, the opinion of most of the secondary school appraisers was that this knowledge would have been learned in pre-Grade Thirteen French courses.

In Part Three of the test the student read a prose passage aloud after having a minute to study its contents. Table A3.14 is a summary report of appraisers' estimates of the percentage of students who could give an acceptable reading of this passage upon entry to and upon successful completion of a Grade Thirteen or a

first year university course in French. Seventy-five per cent of the secondary school appraisers felt that 75 per cent or more of students at the interface, that is, students who had successfully completed a Grade Thirteen course, should be able to read the passage in acceptable fashion. Only half the university appraisers held the same opinion of the abilities of students entering first year university courses.

Other opinions about Part Three of the Speaking Test were solicited through three additional questions (see Table A3.15). The first of these concerned the estimated difficulty of the reading passage. A majority of the appraisers, teaching either Grade Thirteen or first year university courses, indicated that the level of difficulty was about right for students at the interface. A second question asked whether the skills needed to read the passage were taught in Grade Thirteen or in first year university. Responses, summarized in Table A3.15, indicate that this type of skill did receive some degree of emphasis in the courses taught by most of the appraisers. Finally, the criteria for scoring a rendition of the passage were presented for consideration (see Question 3, Table A3.15). Most of the secondary school instructors felt that other criteria of equal or greater importance should be used in scoring. On the other hand, most of the university appraisers said that phrasing and pronunciation were adequate in themselves. Those appraisers who felt that other criteria should be used, had the following suggestions: intonation, dramatization or expression, speed, stress and use of liaison.

Part Four of the Speaking Test consisted of two exercises in fluency. In one case the student was to tell a story about the events depicted in a series of cartoons; in the other case, the student was to describe what he/she thought had happened prior to, during, and after the event depicted in a single picture. Appraisers were asked to estimate the percentage of students who could give acceptable responses to each of these tasks. The tabulation of these estimates, which appears in Table A3.16, indicates that more than one-half of the appraisers expressed the

opinion that more than 75 per cent of students at the interface should have the knowledge required to give an acceptable response to the tasks in Part Four of the test. (The large number of secondary school appraisers--approximately one-third--who failed to respond to the two questions concerning the ability of students entering Grade Thirteen courses to give an acceptable response to the tasks in Part Four of the test is noted. When the Test Appraisal Inventories were given to the appraisers, the response space was inadvertently omitted and very few respondents undertook to reply without an indicated space.)

The tabulations of responses to four other questions about Part Four of the Speaking Test appear in Table A3.17. From this table, it is apparent that fluency of the kind required to do this part of the test received some emphasis in the courses of a majority of the secondary school appraisers and all of the university appraisers who responded. It is additionally noted that most appraisers felt that the tasks in Part Four of the test were at a reasonable level of difficulty. Finally, it can be seen that a sizeable number of the appraisers felt that the criteria of total number of clauses, pronunciation, vocabulary, and grammar were not sufficient for judging students' responses to the fluency tasks. Suggested as additional or alternative criteria were the following: ease of expression or fluency, rhythm, speed, clarity, coherence, originality, humor, sentence complexity, and idiomatic proficiency.

Finally, the appraisers were asked whether they used tests like the Speaking Test in evaluating the fluency with which their own students spoke French. The results presented in Table A3.18 indicate that a majority of both the secondary school and the university appraisers made some use of this kind of test, although a sizeable fraction did not.

The following conclusions about the Speaking Test seem to be in order:

- (a) The first two parts of the test assessed skills that students at the interface should have, but apparently most students would have acquired these skills before Grade Thirteen. These parts of the test were not really geared to the learning students did in the Grade Thirteen year.
- (b) The oral reading and fluency parts of the test assessed skills that most instructors expected most students at the interface to have. The tasks contained in these parts of the test seemed to be at a reasonable level of difficulty. For these reasons, as well as for the reason that oral reading and fluency skills received some attention in Grade Thirteen and first year university courses, the third and fourth parts of the test seemed appropriate for use in assessing the French skills of students at the interface.
- (c) Other criteria in addition to those used to assess the oral reading and fluency exercises were judged to be important by the appraisers. To this extent, the assessment provided by these parts of the test will be narrower than most appraisers would like.

2.5 Concluding Statement

The overall impression conveyed by the appraisals of the Tests of French as a Second Language is one of qualified acceptance. Some parts of the tests seemed inappropriate for assessing students at the interface because they focussed on knowledge that most secondary instructors expected would have been learned prior to Grade Thirteen. On the other hand, several parts of the tests contained exercises judged to be at an appropriate level for interface students and also to tap skills that are important.

3. TECHNICAL ISSUES

3.1 Scoring the Multiple-choice Tests

The items in both the Reading Test and the Listening Test were of the multiple-choice type, each item having four response options. In the instructions for working these tests, students were informed that a correction for guessing would be employed in scoring. The standard correction for four-choice items is to assign each incorrect answer a weight of $-1/3$. In this scoring scheme, correct answers are weighted 1 and omitted questions are ignored.

3.2 Scoring the Writing and Speaking Tests

These tests posed questions to which the student responded, in one case in written form and in the other case in spoken form. These responses were then scored by one or more markers using procedures that could be applied relatively objectively. These procedures were developed for use in the IEA study of achievement in French (Carroll, 1975).

Each marking of a Writing Test yielded four numbers. The first of these, referred to hereafter as Total Writing I, was the number of correct answers to questions about grammar and about verbs and modifiers. These questions were contained in the first part of the Writing Test. Responses to these questions consisted of single words, each of which was scored 0 (wrong) or 1 (correct). As 42 different responses were called for, scores on this part of the test could range from 0 to 42.

The remaining three numbers were assigned to the composition written in response to the directed composition exercise contained in part two of the Writing Test. These numbers Carroll (1975, p. 78) defined as follows:

a - the number of intelligible clauses;

b - the number of clauses which are grammatically correct;

c - the number of clauses which are correct in vocabulary.

From these numbers, Carroll derived two others:

d - the proportion of clauses

Writing Quantity = $40d + 30e$;

e - the proportion of clauses correct in vocabulary, i.e.
 $e = c/a$.

From these numbers, Carroll suggested the computation of the following three scores:

Writing Quantity = $3b + 2c$;

Writing Quality = $40d + 30e$;

Total Writing II = Writing Quantity and Writing Quality

The Speaking Test yielded a number of different scores. The first was based on performance on the pronunciation exercise. There were 16 sentences in this part of the test, each of which was to be spoken by the student. Contained in the 16 sentences were 29 critical phonemes or phonological features of French. The student's rendition of each phoneme or phonological feature was scored 0 (unacceptable) or 1 (acceptable). Consequently, the range of possible scores on the pronunciation part of the test was 0 to 29.

The second score was for "structural control". The task a student performed in this part of the test could be described as follows: he heard a tape-recorded voice ask a question about each of 10 cartoons and he was to respond appropriately to the

question within the context set by the cartoon. Each response was scored on a scale from 0 (low) to 4 (high). The points on the scale Carroll (1975, p. 72) defined as follows:

- 0 - no response or a response of "yes" or "no" or an inappropriate or unintelligible response or a response of "I don't know";
- 1 - an accumulation of serious errors in the response or one serious error accompanied by one or several minor errors;
- 2 - one serious error or an accumulation of minor errors;
- 3 - one or two minor errors;
- 4 - completely correct.

The judgments required to score these responses involved distinguishing serious errors from minor ones; following the procedures developed for the IEA study, no list of serious and minor errors was supplied the marker, but it was hoped that this distinction was made consistently by the marker throughout the marking process. The scale for a score on this part of the test ranged from 0 to 40.

The third part of the Speaking Test was oral reading. The paragraph that students were asked to read contained 15 sentences, 165 words of "relatively simple French prose" (Carroll, 1975, p. 72). Carroll went on to state, "Scoring was based on a series of 24 points concerning particulars of pronunciation (including stress and intonation); a score of 0 or 1 was given for each point depending upon its acceptability. Thus, the total scores could range from 0 to 24" (Carroll, 1972, p. 72).

The final part of the Speaking Test contained two fluency exercises. In one exercise, the student was to describe what had happened in a series of three cartoon pictures. In the other exercise, the student was to look at a cartoon scene and "relate the probable action that took place before the events depicted in the picture, the actual events, and the probable outcome of the situation". Responses to these exercises were scored for the following characteristics (Carroll, 1975, p. 75):

- a - the number of intelligible clauses in the response;
- b - the number of different grammatical structures represented in the response;
- c - the number of clauses correct in structure;
- d - the number of clauses correct in morphology (inflection);
- e - the number of clauses correct in vocabulary;
- f - the number of clauses correct in pronunciation.

Once these numbers were determined, the marker made a global rating of the two fluency responses on the following scale (Carroll, 1975, p. 75):

- 0 - very bad or no response;
- 1 - poor but passable;
- 2 - satisfactory;
- 3 - good;
- 4 - very good.

Carroll (1975, p. 75) suggested the use of the following formulas:

$$\text{Fluency Quantity Score} = 2b + 2c + d + 2e + 2f;$$

$$\text{Fluency Quality Score} = (6c + 3d + 6e + 6f)/a;$$

$$\text{Total Fluency} = \text{Fluency Quantity Score} + \text{Fluency Quality Score};$$

where a, b, c, d, e and f are as defined above, and scores are set to zero when a = 0.

Carroll also suggested computing a total score on the Speaking Test (1975, p. 76):

$$\text{Total Speaking Score} = \text{Total Fluency Score} + 2(p + s + o),$$

where p = pronunciation score

s = structural control score

o = oral reading score

3.3 Difficulty of the Reading and Listening Tests

The measure of difficulty that was computed for each item in these tests was the average across schools of the percentage of students within each school who answered the item correctly. This measure is inversely related to difficulty in that the larger the percentage of correct responses, the easier the item.

The difficulty indices for the items in the Reading and Listening Tests are reported in Table A3.19. Distributions of these difficulty indices and the means and standard deviations of the distributions are given in Table A3.20. These results indicate that, on average, approximately 65 per cent of the Grade

Thirteen students writing these tests were able to respond correctly to the reading and listening items. The variation of difficulty indices is substantially larger for the Reading Test than for the Listening Test, but in each case the range of indices is from over 90 to less than 30. This degree of variation in item difficulty is not atypical for tests designed to spread out the members of the population being assessed from relatively poor performance to relatively good performance on the test.

3.4 Speededness of the Reading and Listening Tests

Evidence of speededness was sought in the percentage of students who failed to respond to an item. An arbitrary distinction was made, in compiling this evidence, between an omitted item and an item that had not been reached. A failure to respond to an item was taken as an indication that the item was not reached if the student failed to respond to all the items that followed the item in question in the test. Otherwise the failure to respond was classed as an omission.

The percentages of failures to respond that were classed "omit" and "not reached" for each item in the Reading Test are reported in Table A3.21. Because the test was divided into two separately-timed sections, one would expect to see not reached percentages different from zero for items toward the end of the first part of the test. Unfortunately, the fact that the Reading Test was in two separate parts was ignored in this analysis. Consequently evidence of speededness in the first part of the test had to be sought in the percentage of omissions.

If the first part of the Reading Test was speeded, the percentages of omissions should grow systematically larger from one item to the next. What happens, in fact, is that the percentages of omissions vary widely from item to item over the 17 items in this part of the Reading Test (see Table A3.21). Systematic growth in the percentage of omissions, if it occurs at

all, is restricted to items 15, 16 and 17, and given that other items show even larger percentages of omissions than item 17, it is possible that the trend in the percentages of omissions for items 15, 16 and 17 is due simply to the fact that the corresponding number of students didn't have the requisite knowledge to answer these questions. Suppose, however, that all the omissions of items 15, 16 and 17 were caused by test speededness. One rule of thumb for assessing speededness is to ascertain whether less than 100 per cent of examinees completed 75 per cent of the test and whether less than 80 per cent of examinees completed all of the test. If this rule is applied to the percentage of omissions for items 15, 16 and 17 of the Reading Test on the assumption that these percentages reflect the fact that examinees were unable for want of time to complete the first part of the test, then it must be concluded that the test was speeded to some extent. But the degree of speeding does not appear to be serious.

The percentages of failures to respond classed as "not reached" for items 31-39 of the Reading Test provide a fair assessment of the speededness of the second part of this test. By the rule of thumb cited in the preceding paragraph, this part of the test was only minimally speeded, if at all.

The Listening Test could not be assessed for speededness in the way that the Reading Test was, for it was administered in a different way. In the Listening Test, students heard a question spoken to them by a tape-recorded voice and then they had a brief interval of time to respond. This interval might have been too short for some students, in which case the percentage of persons not responding to an item would be larger than it would have been if more time had been allowed for responding. But without any evidence on what the percentage of failures to respond would be if a longer response time had been given, the percentages of omissions reported in Table A3.21 are better interpreted as evidence of lack of knowledge on the part of a number of the students taking the test than as evidence of speededness.

In summary, the figures reported in Table A3.21 do not create cause for concern about speededness as a factor affecting student performance on either the Reading Test or the Listening Test.

3.5 Scores on the Writing Test

This instrument was included in the study because it was felt the study would be subject to criticism if writing were not assessed. But it was necessary to limit the number of students taking the Writing Test for the very practical reason that it was time consuming, and therefore expensive to score. In all, only 54 students wrote the test.

The 54 test papers were scored as follows: They, along with a set of scoring instructions, were given to each of two markers. These individuals scored the first five test papers independently, then met to compare results and discuss differences. Finally, they independently marked the remaining papers.

The means, standard deviations and low-high values of the distributions of scores assigned by the two markers are presented in Table A3.22. The four scores are those defined previously in the section on scoring the Writing Test. It is clear from these figures alone that the markers assigned very similar scores to the same test paper. A further indication of the degree of agreement the markers achieved is provided by the correlation coefficients reported in Table A3.23. Agreement is best for Total Writing I, a result not surprising in view of the fact that this part of the test was marked very objectively. The other scores called for greater exercise of judgement, especially in distinguishing an error in grammar from one of vocabulary. Even so, the correlation coefficients are acceptably high for this kind of marking task.

The corresponding scores assigned by the two markers to the same test paper were averaged. These averages were then used to compute Writing Quantity, Writing Quality and Total Writing II scores. (See the definition of these scores given earlier in the report.)

Some important information about the Writing Test is contained in the means and standard deviations of the distributions of all the different scores on the test; these are reported in Table A3.24. Total Writing I, the score for Part One of the Writing Test, has a fixed scale; the possible range of scores is 0-42. The 54 students from across Ontario scored, on average, at approximately the mid-point on the scale. This is an ideal result for a constructed response test designed to spread students out in terms of their performance in French writing. The standard deviation of 7.5 for these scores indicates that the students were in fact spread widely over the scale in terms of their performance on the test. Most scores on Part Two of the Writing Test do not have a fixed scale or fixed upper limit. The exceptions are the two proportion scores, which obviously range from 0 to 1. The relatively large standard deviations of the various Part Two scores indicate substantial variability about the means in the performance of the students. In the sense that it spread students out in terms of their performance, this part of the test was also functioning effectively.

The Writing Test was used in an international study of achievement in French as a second language (Carroll, 1975). Results from this study provide a rough basis against which to assess some of the means and standard deviations reported in Table A3.24. Any comparison of the results achieved in the two studies is rough because different individuals were involved in the scoring; consequently, it is impossible to say whether the scoring criteria were applied in exactly the same way in both studies. Nevertheless, it is important to note that where comparisons are possible, the means and standard deviations reported in Table A3.24 fall within the range of means and standard deviations reported for the different countries

participating in the international study (see Carroll, 1975, pp. 164-165).

It is possible to obtain a crude estimate of the reliability of Total Writing I scores by using the mean and standard deviation reported in Table A3.24 and Kuder-Richardson formula 21 (Lord and Novick, 1968, p. 91). This estimate is crude because it is only a lower bound, which is to say, the scores are very probably more reliable than the estimate suggests. Even so, the reliability, as estimated by this method, is 0.83, a result that is acceptably high for this kind of test and within the range of estimates observed by Carroll (1975, p. 96). The associated standard error of measurement of 3.1 is, again, well within the range of values observed by Carroll for this test.

The data generated by the directed composition exercise in part two of the Writing Test cannot be used to estimate reliability coefficients and associated standard errors of measurement for scores on this part of the test.

The question arises as to whether or not students are placed at the same relative positions along the different scales of the Writing Test. One way to answer this question is to compute coefficients of correlation among the various scores on the test. These correlation coefficients are reported in Table A3.25. Because all the different scores on Part Two of the test are based on the same exercise, these scores are not independent and some degree of correlation is to be expected. There are, nevertheless, several points to be made on the basis of the results given in Table A3.25.

- (a) Total Writing I is independent of all the other scores. Its highest correlation with another score is .66 with Total Writing II. This indicates that the two parts of the test did allocate students to somewhat different relative positions along the corresponding scales of the test. In other words, the two tests appear to measure

two somewhat different, though related, French writing abilities.

- (b) The correlation of 0.76 between Writing Quantity and Writing Quality is very near the correlations of 0.71 and 0.77 reported by Carroll (1975, p. 79 and p. 107) for two different groups.
- (c) A further comparison with Carroll's international study is possible using the correlations of Total Writing I with each of Writing Quantity and Writing Quality. In the present study, these correlations are 0.64 and 0.59 respectively; Carroll reports figures of 0.59 and 0.60.
- (d) Total Writing II is very closely related to Writing Quantity and its correlation with this variable is much higher than it is with Writing Quality. This indicates that the majority of variation in Total Writing II is accounted for by variance in Writing Quantity. To preserve the unique information contained in Writing Quality scores, both Writing Quantity and Writing Quality scores should be included in analyses of performance on Part Two of the Writing Test and the single score Total Writing II should be ignored.

In summary, all indications are that the Writing Test functioned in an acceptable manner in Project II. Moreover, the results achieved using this test are about what would be expected, given the way the test performed in a previous international study in which it was used.

3.6 Scores on the Speaking Test

This instrument was also included in the study to avoid the criticism that might have been advanced against the study if speaking skills had been totally neglected. Nevertheless, it was

necessary, because of the amount of time required to administer and, subsequently, to score the test, to limit the number of students assigned to take it. In all, 58 students attempted at least part of the test, and 50 students completed the whole test. The results reported here are for the 50 students who completed the test.

The 58 tape-recorded responses were scored by a single marker. This person had had previous experience scoring the Speaking Test.

The means, standard deviations and minimum and maximum scores of the distributions of scores on the Speaking Test are reported in Table A3.26. These results should be interpreted, bearing in mind that Pronunciation, Structural Control, Oral Reading and Global Rating of Fluency have fixed score scales with maximum possible scores of 29, 40, 24 and 4, respectively. All the other score scales do not have maximum possible values. In general, the test seems to have produced reasonable results. The range of scores is narrower than might have been desired on the Pronunciation and Oral Reading scales. But when these results are compared with those obtained in an international study (Carroll, 1975, pp. 85, 87, 88, 164, 165 and 224), it is apparent that the test was functioning in this study in much the same way that it had in the international study.

Crude estimates of reliability were obtained for Pronunciation, Structural Control and Oral Reading scores by substituting their respective means and standard deviations in Kuder-Richardson formula 21 (Lord and Novick, 1968, p. 91). The estimate of reliability provided by this formula is crude because it is known to be a worse lower-bound estimate than that provided by other means. Unfortunately, the data required to apply another means of estimating reliability were not available for these scores. Besides being crude, it is doubtful that KR-21 is applicable to the Structural Control scores. KR-21 was derived for application to tests comprised of dichotomously scored items. The Pronunciation and Oral Reading parts of the Speaking Test may

be regarded as composed of dichotomously scored items; the marker focussed on 29 and 24 linguistic features in these two parts of the test and scored each of them 0 for incorrect or unacceptable and 1 for correct or acceptable. The Structural Control part of the test, on the other hand, was composed of ten items and the student's response to each item was rated on a scale from 0 to 4 (see previous discussion of scoring). When KR-21 is used to estimate the reliability of Structural Control scores, each score on the test is being treated as if it arose as a result of 40 separate dichotomous judgments of correct or incorrect, not 10 ratings on a 0-4 scale. Nevertheless, Carroll (1975) used KR-21 with Structural Control and hence it was applied here for comparative purposes.

The reliability estimates for Pronunciation, Structural Control and Oral Reading were 0.52, 0.88 and 0.63 respectively. Associated with these figures are estimates of the standard error of measurement: 2.4, 2.9 and 2.2. The estimates of reliability for the Pronunciation and Oral Reading scores are disappointingly low when judged in absolute terms and the associated standard errors of measurement are large relative to the standard deviations of scores on these parts of the Listening Test. It is true, nevertheless, that the reliability estimates and the standard errors of measurement of these scores and of the scores on Structural Control are well within the range of values reported by Carroll (1975, p. 94).

Estimates of reliability for the counts that were made in scoring the fluency tasks were achieved in a different way. This part of the test was comprised of two exercises. Each exercise was scored separately. The estimates of reliability were based on correlations between corresponding scores on the two exercises. The results are reported in Table A3.27. These figures are very close to values reported by Carroll (1975, p. 95) for the same scores. (Elsewhere, corresponding scores in both fluency exercises are summed to yield a single score of each type. These summed scores generated the results reported in Table A3.26).

Because of the nature of the data that were collected and because all the Speaking Tests were scored by just one person, it was not possible to estimate either the reliability of the Global Rating of Fluency or the degree of rater consistency that it is possible to achieve in the Speaking Test. Carroll reported (1975, pp. 94 and 95) estimates of rater consistency that are acceptably high for this type of test. There is no reason to believe that the trained marker used in the present study would be markedly less consistent than the markers who scored the Speaking Test in the international study.

A further study of the Speaking Test was made by intercorrelating all the different scores on the test. The resulting correlations are reported in Table A3.28. Carroll reported correlations between some of these same variables (1975, pp. 75, 76, 107) for data collected in the international study. The correlations he reported among Pronunciation, Structural Control, and Oral Reading, and the correlations he reported between these three variables and Fluency Quantity and Fluency Quality, are somewhat lower than the correlations observed in the present case. But the correlations Carroll reported among Fluency Quantity, Fluency Quality, and Total Speaking Test and those between either Total Fluency or Total Speaking Test and Pronunciation, Structural Control and Oral Reading are very close to the corresponding correlations observed in this study.

One other conclusion can be advanced on the basis of the results presented in Table A3.28. The correlation between Total Fluency and Fluency Quantity is almost perfect whereas the correlation between Fluency Quality and Total Fluency is very much less (0.77). This indicates that the Total Fluency and Fluency Quantity are redundant and that performance on the fluency exercises of the Speaking Test is best described when Fluency Quantity and Fluency Quality are included as separate variables.

From the preceding summary of evidence, it appears that the Speaking Test generated results in the present study that are both

reasonable and very much in line with what would be expected on the basis of past experience with the test.

3.7 Reading Test and Listening Test--Information on Discriminations, Reliability and Distribution of Scores

Discrimination is an important characteristic to note about the items of tests designed to spread students across the range of possible scores. The quality of discrimination can be loosely defined as the extent to which student performance on an item is correlated with overall performance on the test. The biserial coefficient of correlation between item scores and total test scores is frequently used as an index of item discrimination. By one rule of thumb, biserial coefficients of correlation greater than 0.3 can be considered acceptable.

The indices of item discrimination for the items in the Reading Test and the Listening Test are in Table A3.29. These indices were computed from the test responses of all the students who wrote the Reading Test and the Listening Test. These students were all enrolled in Grade Thirteen courses in French in the participating schools of the study. From the results reported in Table A3.29, it can be seen that all items in both tests, except for two in the Reading Test--item 18 and item 21--had indices of discrimination in excess of 0.3. The two items with relatively low discrimination indices were very easy (see Table A3.19). This cannot be taken as evidence, however, that all very easy items have low discrimination indices; item 1 in the Reading Test was also very easy for these students but it has an acceptably high discrimination index. The obvious conclusion from these results is that acceptable levels of item discrimination were realized for the vast majority of items in both tests.

Distribution statistics and information on reliability and error of measurement are reported in Table A3.30. It can be seen from the means reported in this table that both the Reading Test and the Listening Test were relatively easy for Ontario

SSHGD-level students. An ideal mean for the distribution of corrected-for-guessing scores on a test designed to spread students over the range of possible scores would be approximately one-half the number of questions on the test. The means on both the Reading Test and the Listening Test exceed the ideal.

The distributions of scores on the tests (not shown) were differently shaped. Despite its relatively high mean, the distribution of corrected-for-guessing scores on the Reading Test was approximately bell-shaped. Only three persons got a perfect score on the test and only one got a negative score. (A negative score reflects performance at a level below that which would be expected were all questions to be answered by guessing.) The distribution of scores on the Listening Test was, in contrast, negatively skewed. There was a preponderance of high scores, reflected in part by the fact that 14 students achieved perfect scores on the test. But the distribution tailed off into the region of negative scores--5 students performed less well on the test than they might have expected to do by guessing.

The coefficients of reliability for both tests are acceptably high for the purposes of the study. These purposes are to compare the performance of different groups of students and to predict school marks using scores on the tests.

Both the Reading Test and the Listening Test were used in an international study of achievement in French (Carroll, 1975). Carroll reported (see p. 93) means, standard deviations, reliability coefficients and standard errors of measurement for these tests for each of several different countries. Unfortunately, the figures Carroll reported in his discussion of the technical qualities of the tests are not directly comparable with those obtained in the present application of the tests. There are two reasons for this: (i) Both the Reading Test and the Listening Test were modified somewhat for administration to Ontario students. The wording of the Reading Test was changed in 14 places; the last six items of the Listening Test used in the international study were deleted in adapting it for use in

Ontario. In addition, the procedure for administering five items of the Listening Test was changed; instead of hearing the material for items 30 to 34 of the test spoken once, as in the international study, students in the present study heard the material spoken twice. All these changes were made in an effort to render the tests more appropriate for use with Ontario students. (ii) Carroll did not employ a correction-for-guessing in his analysis of test reliability, although the students were told a correction would be applied.

Because of the differences in the testing methods used in the two studies, it is difficult to predict how all the Ontario results would change if the method of the international study had been strictly followed. It is probable that the mean reported in Table A3.20 for the Listening Test is lower than it would have been had the method of testing used in the international study been followed exactly, but the effect of the differences in method on the standard deviation, reliability coefficient and standard error of measurement is difficult to predict. Even so, the results obtained in the present study fall within the range of results reported by Carroll. In comparison with the results achieved in the international study, the Reading Test and the Listening Test both seem to have functioned well in the present application.

REFERENCES

- Carroll, J.B. The teaching of French as a foreign language in eight countries. Toronto: John Wiley and Sons, 1975.
- Hoyt, C. Test reliability estimated by analysis of variance. Psychometrika, 1941, 6, pp. 153-160.
- Lord, F.M. and Novick, M.R. Statistical theories of mental test scores. Toronto: Addison-Wesley, 1968.

TABLE A3.1

Number of Respondents Choosing Each Category of Response in
Appraising the Items in Part A of the Reading Test

Test Item Number	Secondary School Appraisers ^a (N = 64)					University Appraisers (N ^a = 18)						
	Response Categories					Response Categories						
	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C1</u>	<u>C2</u>	<u>C3</u>	<u>C4</u>	<u>D</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C</u>
1	47	17	-	-	-	-	-	-	15	1	1	-
2	29	21	8	1	-	4	-	-	9	5	3	1
3	40	23	1	-	-	-	-	-	11	5	1	-
4	40	23	1	-	-	-	-	-	10	8	-	-
5	47	12	4	-	-	-	1	-	11	3	4	-
6	6	17	26	1	1	2	6	4	2	7	8	1
7	15	17	15	2	1	6	4	4	3	8	5	2
8	5	5	27	1	3	5	8	9	1	5	11	1
9	9	15	16	1	5	4	7	6	3	4	9	2
10	9	12	26	0	2	10	2	3	3	9	5	1
11	6	9	20	2	3	3	7	14	2	7	9	-
12	17	23	15	0	2	4	2	1	7	9	2	-
13	17	19	18	0	5	2	2	1	5	7	6	-
14	6	12	12	0	4	5	11	12	4	5	7	2
15	8	9	20	0	0	7	7	13	2	7	8	1
16	3	10	24	1	3	3	7	13	2	8	4	4
17	3	4	17	2	3	4	11	19	2	3	12	1

TABLE A3.2

A Tabulation of Judgements of Difficulty and Appropriateness
on each Passage Contained in Part B of the Reading Test

Passage Number	Response Category	Assessment of Difficulty		Assessment of Appropriateness ^a	
		Secondary School Appraisers (N = 64)	University Appraisers (N = 18)	Secondary School Appraisers (N = 64)	University Appraisers (N = 18)
1	Too easy	23	5	Yes	4
	Somewhat easy	33	7		
	About right	8	5	No	59
	Somewhat difficult	-	1		
	Too difficult	-	-		
2	Too easy	24	4	Yes	5
	Somewhat easy	19	7		
	About right	11	7	No	58
	Somewhat difficult	-	-		
	Too difficult	-	-		
3	Too easy	3	-	Yes	5
	Somewhat easy	18	2		
	About right	33	10	No	58
	Somewhat difficult	10	5		
	Too difficult	-	1		
4	Too easy	5	1	Yes	3
	Somewhat easy	22	4		
	About right	31	10	No	60
	Somewhat difficult	6	2		
	Too difficult	-	1		

TABLE A3.2 (continued)

Passage Number	Response Category	Assessment of Difficulty		Assessment of Appropriateness ^a	
		Secondary School Appraisers (N = 64)	University Appraisers (N = 18)	Secondary School Appraisers (N = 64)	University Appraisers (N = 18)
5	Too easy	2	-	Yes 15	2
	Somewhat easy	5	1		
	About right	28	5	No 47	14
	Somewhat difficult	23	8		
	Too difficult	6	4		
6	Too easy	1	-	Yes 6	1
	Somewhat easy	4	1		
	About right	30	8	No 56	15
	Somewhat difficult	26	4		
	Too difficult	3	5		

Note: Omitted responses account for discrepancies between the number of tabulated responses and the number of appraisers.

^aAppraisers were asked the following question: "Is the passage inappropriate for reasons other than difficulty?"

TABLE A3.3

A Tabulation of Estimates of the Percentage of Students
Entering and Completing French Courses at the Grade 13 and
University Levels Who Should Possess Each of Three
Reading Abilities

Question	Response	Level of Courses			
		Grade 13 (N=84)		First Year University (N=18)	
		On Entry	On Successful Completion	On Entry	On Successful Completion
1. How many students should be able to read passages like those in the test for literal understanding?	All	15	32	9	12
	More than 75%	21	24	5	5
	51% to 75%	14	5	-	-
	26% to 50%	8	2	4	1
	Less than 25%	5	-	-	-
	None	-	1	-	-
2. How many students should be able to identify the main idea or purpose of passages like those in the test?	All	16	39	8	12
	More than 75%	22	23	3	5
	51% to 75%	17	1	4	-
	26% to 50%	6	1	2	1
	Less than 26%	2	-	1	-
	None	-	-	-	-

TABLE A3.3 (continued)

<u>Question</u>	<u>Response</u>	<u>Level of Courses</u>			
		<u>Grade 13</u> (N=84)	<u>First</u> <u>Entry</u>	<u>Year University</u> (N=18)	<u>On Successful</u> <u>Completion</u>
3. How many students should be able to draw inferences from and see implications in the ideas expressed in passages like those in the test?	All	5	1	4	
	More than 75%	15	7	6	
	51% to 75%	19	3	4	
	26% to 50%	12	4	3	
	Less than 26%	12	2	1	
	None	-	1	-	

Note: Omitted responses account for discrepancies between the number of tabulated responses and the number of appraisers.

TABLE A3.4

A Tabulation of Responses to a Question
Concerning the Degree of Emphasis Given
Each of Three Reading Skills in
Courses at the Interface

<u>Reading Skill</u>	<u>Judged Degree of Emphasis</u>	<u>Secondary School Appraisers (N=84)</u>	<u>University Appraisers (N=18)</u>
1. Reading for literal understanding	Heavy emphasis Moderately heavy emphasis Light emphasis Emphasis on individual remediation No emphasis	23 22 16 3 -	8 4 3 2 1
2. Reading to identify the main idea or purpose	Heavy emphasis Moderately heavy Light emphasis Emphasis on individual remediation No emphasis	31 27 5 1 -	7 10 1 - -
3. Reading to draw inferences and see implications	Heavy emphasis Moderately heavy emphasis Light emphasis Emphasis on individual remediation No emphasis	17 32 12 2 1	7 6 4 1 -

TABLE A3.5

A Tabulation of Responses to Three General Questions
About the Reading Test

<u>Question</u>	<u>Response</u>	<u>Secondary School</u> <u>Appraisers (N=64)</u>		<u>University</u> <u>Appraisers (N=18)</u>	
		Yes	No	Yes	No
1. Are there important reading skills that are not assessed by the <u>Reading Test</u> but which should be?	Yes	14		5	
	No	48		12	
2. Is the multiple-choice format a reasonable method of assessing at least those reading skills covered by the <u>Reading Test</u> ?	Yes	25		8	
	Yes (Qualified)	35		9	
	No	3		1	
3. Do you use this kind of test in evaluating student skills in reading?	Yes, frequently	3		7 ^a	
	Yes, occasionally	39			
	No	21		11	

Note: Omitted responses account for discrepancies between the number of tabulated responses and the number of appraisers.

^aThe appraisal inventory for university instructors offered only one, unqualified response of Yes.

TABLE A3.6

Number of Respondents Choosing each Category of Response in Appraising the Items of the <u>Listening Test</u>														
Test Item Number	Secondary School Appraisers ^a (N = 64)					University Appraisers (N ^a = 18)								
	A1	A2	B	C1	.	C2	C3	C4	D	A1	A2	.	B	C
1	50	5	-	-	.	-	-	-	-	15	3	.	-	-
2	37	16	2	-	.	-	-	-	-	10	8	.	-	-
3	43	10	-	-	.	-	1	1	-	13	5	.	-	-
4	48	6	-	-	.	-	-	-	-	14	4	.	-	-
5	43	11	-	-	.	-	-	-	-	9	9	.	-	-
6	20	25	9	-	.	-	-	1	-	5	8	.	3	2
7	42	12	1	-	.	-	-	-	-	10	7	.	-	1
8	48	7	-	-	.	-	-	-	-	12	5	.	-	1
9	45	10	-	-	.	-	-	-	-	11	7	.	-	-
10	22	27	3	1	.	-	-	1	-	8	7	.	3	-
11	33	18	3	-	.	-	-	1	-	7	8	.	2	1
12	18	21	14	-	.	1	-	1	-	7	7	.	4	-
13	16	24	11	-	.	1	1	2	-	4	8	.	6	-
14	25	16	10	2	.	-	-	1	-	6	5	.	5	2
15	10	5	18	2	.	2	2	4	12	4	3	.	9	2
16	21	23	8	-	.	1	-	2	-	8	3	.	6	1
17	38	14	1	-	.	-	-	1	-	11	5	.	-	1
18	32	19	2	-	.	-	1	-	-	6	9	.	1	1
19	30	21	2	-	.	-	1	-	-	9	7	.	-	1
20	14	16	13	1	.	1	5	2	2	2	7	.	6	2
21	18	17	11	1	.	-	2	1	4	3	7	.	7	-
22	22	20	9	-	.	-	-	1	2	3	11	.	2	1
23	34	16	3	-	.	-	-	1	-	6	10	.	1	-
24	11	15	12	-	.	-	3	6	7	2	5	.	6	4

TABLE A3.6 (continued)

Test Item Number	Secondary School Appraisers (N = 64)						University Appraisers (N ^a = 18)					
	Response Categories						Response Categories					
	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C1</u>	<u>C2</u>	<u>C3</u>	<u>C4</u>	<u>D</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C</u>
25	14	21	18	-	1	-	-	-	8	6	3	1
26	26	22	4	1	-	1	-	-	9	6	2	1
27	9	9	23	1	2	4	2	4	4	6	7	1
28	11	17	18	1	-	2	2	2	4	6	8	-
29	6	9	21	1	1	1	7	6	6	6	5	1
30	29	17	3	1	2	-	-	1	6	9	2	1
31	32	14	4	1	2	-	-	-	7	7	2	2
32	22	21	4	-	4	1	1	-	7	6	2	3
33	20	23	5	1	1	1	1	-	6	5	3	4
34	22	21	6	-	1	1	1	-	6	7	2	3

Note: Omitted responses account for discrepancies between the number of tabulated responses and the number of appraisers.

^aResponse Key: Secondary School

University

- | | |
|---|-------------------------------|
| A1 - Old knowledge, no review | A1 - Old knowledge, no review |
| A2 - Old knowledge, review | A2 - Old knowledge, review |
| B - New knowledge, all learn | B - New knowledge, all learn |
| C1 - New knowledge, more than 75% but not all learn | C - Other |
| C2 - New knowledge, 51% to 75% learn | |
| C3 - New knowledge, 26% to 50% learn | |
| C4 - New knowledge, 1% to 25% learn | |
| D - New knowledge, none learn | |

TABLE A3.7

A Tabulation of Responses to Four General Questions
About the Listening Test

<u>Question</u>	<u>Response</u>	Secondary School	University
		Appraisals (N=64)	Appraisals (N=18)
1. What emphasis do you give to listening in your Grade 13/first- year university French course?	Heavy emphasis	15	9
	Moderately heavy emphasis	22	8
	Light emphasis	19	-
	Emphasis in individual remediation	1	1
	No emphasis	2	-
2. Are there important listening skills that are not assessed by the <u>Listening Test</u> but which should be assessed?	Yes	11	4
	No	41	14

TABLE A3.7 (continued)

<u>Question</u>	<u>Response</u>	Secondary School Appraisals (N=64)	University Appraisals (N=18)
3. Is the use of a tape recorded script in conjunction with a multiple-choice format a reasonable method for assessing at least those listening skills covered by the <u>Listening Test</u> ?	Yes	44	11
	Yes (qualified)	12	7
	No	1	-
4. Do you use this kind of test in evaluating student ability to comprehend spoken French?	Yes, frequently	6	13 ^a
	Yes, occasionally	33	5
	No	18	

Note: Omitted responses account for discrepancies between the number of tabulated responses and the number of appraisers.

^aThe appraisal inventory for university instructors offered only one unqualified response of Yes.

TABLE A3.8

Number of Respondents Choosing each Category of Response in Appraising the Items in Part A of the Writing Test

Test Item Number	Secondary School Appraisers (N = 64)					University Appraisers (N _a = 18)						
	Response Categories					Response Categories						
	A1	A2	B	C1	C2	C3	C4	D	A1	A2	B	C
1	34	22	2	2	-	-	-	-	8	9	1	-
2	29	29	1	-	-	1	-	-	6	10	2	-
3	18	38	3	-	-	-	1	-	3	11	4	-
4	15	35	8	-	-	-	1	-	4	14	-	-
5	17	18	20	-	1	2	1	1	4	11	2	1
6	14	35	8	1	-	-	1	1	4	12	2	-
7	9	42	7	-	-	-	2	-	1	14	3	-
8	11	25	16	2	1	1	1	3	3	7	6	2
9	8	36	12	1	-	1	2	-	3	9	6	-
10	11	26	20	-	-	1	-	2	3	6	8	1
11	12	42	5	-	-	1	-	-	6	12	-	-
12	10	21	27	-	-	1	-	-	5	8	4	1
13	25	34	-	-	-	1	-	-	3	15	-	-
14	12	24	18	-	-	2	-	2	2	8	6	2
15	10	42	7	-	-	-	-	1	3	13	2	-
16	6	42	10	-	2	-	-	-	3	10	5	-
17	18	38	3	1	-	-	-	-	5	12	1	-
18	21	36	2	-	-	-	1	-	3	12	3	-
19	19	39	-	-	1	1	-	-	6	10	2	-
20	31	27	2	-	-	-	-	-	8	9	1	-
21	30	28	1	-	-	-	1	-	7	11	-	-
22	21	36	3	-	-	-	-	-	3	12	3	-
23	20	38	1	-	-	-	1	-	6	10	2	-

TABLE A3.8 (continued)

Test Item Number	Secondary School Appraisers (N = 64) Response Categories					University Appraisers (N _a = 18) Response Categories						
	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C1</u>	<u>C2</u>	<u>C3</u>	<u>C4</u>	<u>D</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C</u>
24	24	33	3	-	-	-	-	-	3	14	.	-
25	30	28	2	-	-	-	-	-	7	10	.	-
26	9	37	11	-	-	1	-	-	4	9	.	-
27	40	18	1	-	-	1	-	-	3	13	.	-
28	33	24	2	-	-	-	1	-	5	12	.	-
29	28	26	6	-	-	-	-	-	6	11	.	-
30	30	28	2	-	-	-	-	-	7	11	.	-
31	18	33	9	-	-	-	-	-	5	10	.	-
32	26	32	2	-	-	-	-	-	7	10	.	-

Note: Omitted responses account for discrepancies between the number of tabulated responses and the number of appraisers.

^aResponse Key: Secondary School

University

- | | |
|---|-------------------------------|
| A1 - Old knowledge, no review | A1 - Old knowledge, no review |
| A2 - Old knowledge, review | A2 - Old knowledge, review |
| B - New knowledge, all learn | B - New knowledge, all learn |
| C1 - New knowledge, more than 75% but not all learn | C - Other |
| C2 - New knowledge, 51% to 75% learn | |
| C3 - New knowledge, 26% to 50% learn | |
| C4 - New knowledge, 1% to 25% learn | |
| D - New knowledge, none learn | |

TABLE A3.9

A Tabulation of Estimates of the Percentage of Students
Entering and Successfully Completing French Courses at the
Grade 13 and University Levels Who Should Be Able to Write an
Acceptable Composition of the Type Required in the Writing Test

Percentage of Students	Level of Course			
	Grade 13 (N=84)		University (N=18)	
	<u>On Entry</u>	<u>On Successful Completion</u>	<u>On Entry</u>	<u>On Successful Completion</u>
All	17	29	6	10
More than 75% but not all	19	17	6	7
51% to 75%	10	11	4	-
26% to 50%	8	3	1	1
One or more, but less than 26%	6	1	1	-
None	-	-	-	-

Note: Omitted responses account for discrepancies between the number of tabulated responses and the number of appraisers.

TABLE A3.10

A Tabulation of Ratings of Emphasis Given to
Developing Student Competence in
Writing French Compositions

<u>Rating</u>	<u>Secondary School Appraisers (N=64)</u>	<u>University Appraisers (N=18)</u>
Heavy Emphasis	10	3
Moderately Heavy Emphasis	24	7
Light Emphasis	24	5
Emphasis in Individual Remediation	1	2
No Emphasis	1	1

TABLE A3.11

A Tabulation of Responses to Three General Questions
About the Composition Part of the Writing Test and
One Question About the Writing Test as a Whole

<u>Question</u>	<u>Response</u>	<u>Secondary School</u> <u>Appraisers</u> <u>(N = 68)</u>		<u>University</u> <u>Appraisers</u> <u>(N = 18)</u>	
		Yes	No	Yes	No
1. Are there other types of writing in French that Grade 13/first-year university students should be expected to do and that are as important as, or even more important than, the type required in Part B of the <u>Writing Test</u> ?	Yes	51		11	
	No	9		6	
2. Was the writing assignment in Part B at a reasonable level of difficulty for students in the Grade 13/entering a first-year university course in French?	Yes	42		14	
	No	16		4	
3. Four criteria will be used to judge the composition: (i) Total number of clauses (ii) Grammar (iii) Extent of vocabulary (iv) Accuracy in the use of vocabulary (not including spelling or use of accents) Are there other criteria of equal or greater importance?	Yes	45		13	
	No	15		4	

TABLE A3.11 (continued)

<u>Question</u>	<u>Response</u>	<u>Secondary School Appraisers (N = 68)</u>	<u>University Appraisers (N = 18)</u>
4. Considering the <u>Writing Test</u> as a whole, do you use this kind of test in evaluating student ability to write French?	Yes, frequently	19	12 ^a
	Yes, occasionally	37	
	No	5	6

Note: Omitted responses account for discrepancies between the number of tabulated responses and the number of appraisers.

^aThe appraisal inventory for university instructors offered only one unqualified response of Yes.

TABLE A3.12

Number of Respondents Choosing Each Category of Response in Appraising the Items in Part I of the Speaking Test

Test Item Number	Secondary School Appraisers (N = 64)					University Appraisers (N _a = 18)						
	Response Categories					Response Categories						
	A1	A2	B	C1	C2	C3	C4	D	A1	A2	B	C
1	37	18	3	-	-	-	-	-	9	8	.	-
2	39	17	1	-	-	1	-	-	8	8	.	-
3	38	17	1	-	1	1	-	-	9	7	.	-
4	38	18	-	-	1	1	-	-	8	9	.	-
5	44	13	1	-	-	-	-	-	9	8	.	-
6	37	18	3	-	-	-	-	-	8	8	.	-
7	23	31	4	-	-	-	-	-	9	8	.	1
8	35	20	2	-	-	-	1	-	7	10	.	-
9	42	15	1	-	-	-	-	-	10	7	.	1
10	36	18	3	-	-	-	-	1	11	6	.	-
11	39	14	5	-	-	-	-	-	9	9	.	-
12	30	22	6	-	-	-	-	-	8	10	.	-
13	39	18	1	-	-	-	-	-	8	9	.	-
14	32	21	4	-	-	-	-	1	8	9	.	-
15	35	20	2	-	-	-	1	-	12	6	.	-
16	35	21	2	-	-	-	-	-	11	7	.	-

TABLE A3.12 (continued)

Note: Omitted responses account for discrepancies between the number of tabulated responses and the number of appraisers.

^a Response Key: Secondary School		University
A1 - Old knowledge, no review	A1 - Old knowledge, no review	
A2 - Old knowledge, review	A2 - Old knowledge, review	
B - New knowledge, all learn	B - New knowledge, all learn	
C1 - New knowledge, more than 75% but not all learn	C - New knowledge, all learn	
C2 - New knowledge, 51% to 75% learn		
C3 - New knowledge, 26% to 50% learn		
C4 - New knowledge, 1% to 25% learn		
D - New knowledge, none learn		

TABLE A3.13

Number of Respondents Choosing Each Category of Responses in Appraising Items in Part II of the Speaking Test

Test Item Number	Secondary School Appraisers (N = 64)					University Appraisers (N _a = 18)						
	Response Categories					Response Categories						
	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C1</u>	<u>C2</u>	<u>C3</u>	<u>C4</u>	<u>D</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C</u>
1	25	30	1	-	-	-	-	-	6	8	4	-
2	45	11	-	-	-	-	-	-	9	7	-	-
3	25	19	7	1	2	2	-	1	4	9	2	3
4	22	17	8	-	1	2	3	3	7	7	3	1
5	43	13	-	-	-	-	-	-	9	7	2	-
6	20	19	12	-	2	1	-	-	8	8	2	-
7	29	18	7	-	1	-	1	-	6	8	3	1
8	30	24	2	-	-	-	-	-	10	8	-	-
9	46	10	-	-	-	-	-	-	9	9	-	-
10	21	22	9	-	1	1	1	1	8	8	2	-

Note: Omitted responses account for discrepancies between the number of tabulated responses and the number of appraisers.

Response Key:		Secondary School	University
A1	- Old knowledge, no review	A1 - Old knowledge, no review	A1 - Old knowledge, no review
A2	- Old knowledge, review	A2 - Old knowledge, review	A2 - Old knowledge, review
B	- New knowledge, all learn	B - New knowledge, all learn	B - New knowledge, all learn
C1	- New knowledge, more than 75% but not all learn	C1 - New knowledge, more than 75% but not all learn	C - Other
C2	- New knowledge, 51% to 75% learn	C2 - New knowledge, 51% to 75% learn	
C3	- New knowledge, 26% to 50% learn	C3 - New knowledge, 26% to 50% learn	
C4	- New knowledge, 1% to 25% learn	C4 - New knowledge, 1% to 25% learn	
D	- New knowledge, none learn	D - New knowledge, none learn	

TABLE A3.14

A Tabulation of Estimates of the Percentage of
 Students Entering and Successfully Completing French
 Courses at the Grade 13 and First-Year University Levels
 Who Should Be Able to Give an Acceptable Reading of the
Passage in Part III of the Speaking Test

Percentage of Students	Level of Courses			
	Grade 13 (N=84)		University (N=18)	
	<u>On Entry</u>	<u>On Successful Completion</u>	<u>On Entry</u>	<u>On Successful Completion</u>
All	17	28	3	7
More than 75% but not all	21	25	6	6
51% to 75%	16	6	7	4
26% to 50%	5	-	1	1
One or more, but less than 26%	-	-	1	-
None	-	-	-	-

Note: Omitted responses account for discrepancies between the number of tabulated responses and the number of appraisers.

TABLE A3.15
 Tabulation of Responses to Three Questions About
 Part III of the Speaking Test

<u>Question</u>	<u>Response</u>	Secondary School Appraisers (N = 64)	University Appraisers (N = 18)
1. What is your estimate of the reading level of this passage for students enrolled in your Grade 13/entering your first-year university French course?	Very easy	2	-
	Somewhat easy	14	2
	About right	37	12
	Somewhat difficult	6	2
	Very difficult	-	2
2. How much emphasis do you give in your teaching at the Grade 13/first-year university level to developing student competence in reading out loud prose passages of this kind?	Heavy emphasis	12	-
	Moderately heavy emphasis	21	7
	Light emphasis	20	9
	Emphasis on individual remediation	6	1
	No emphasis	-	1

TABLE A3.15 (continued)

Question	Response		Secondary School Appraisers (N = 64)	University Appraisers (N = 18)
	Yes	No		
3. Two criteria will be used to evaluate a reading of this passage:				
(i) phrasing				
(ii) pronunciation				
Are there other criteria of equal or greater importance?				
	41	4		
	18	13		

Note: Omitted responses account for discrepancies between the number of tabulated responses and the number of appraisers.

TABLE A3.16

Tabulation of Estimates of the Percentage of Students
 Entering and Successfully Completing French Courses at the
 Grade 13 and First-Year University Levels Who Should Be Able to
 Give an Acceptable Response to the
Fluency Tasks in Part IV of the Speaking Test

<u>Percentage of Students</u>	<u>Level of Course</u>			
	Grade 13 (N=64)		University (N=18)	
	<u>On Entry</u>	<u>On Successful Completion</u>	<u>On Entry</u>	<u>On Successful Completion</u>
<u>First Fluency Task</u>				
All	8	19	4	10
More than 75% but not all	14	24	6	4
51% to 75%	12	11	4	1
26% to 50%	7	4	1	1
One or more, but less than 26%	3	-	2	-
None				
<u>Second Fluency Task</u>				
All	5	22	2	7
More than 75% but not all	12	16	8	6
51% to 75%	13	13	5	2
26% to 50%	7	3	1	1
One or more, but less than 26%	6	3	1	-
None				

Note: Omitted responses account for discrepancies between the number of tabulated responses and the number of appraisers.

TABLE A3.17

A Tabulation of Responses to Four Questions About
Part IV of the Speaking Test

<u>Question</u>	<u>Response</u>	Secondary School	University
		Appraisers (N=64)	Appraisers (N=18)
1. How much emphasis do you give in your teaching of the Grade 13/ first-year university French course to the development of student ability to handle the kind of speaking task found in Part IV of the <u>Speaking Test</u> ?	Heavy emphasis	6	5
	Moderately heavy emphasis	13	9
	Light emphasis	25	3
	Emphasis in individual remediation	2	-
	No emphasis	10	-
2. Was the first fluency task at a reasonable level of difficulty for students in a Grade 13/ entering a first year university French course?	Yes	48	12
	No	9	4
3. Was the second fluency task at a reasonable level of difficulty for students in a Grade 13/ entering a first-year university French course?	Yes	49	15
	No	8	1

TABLE A3.17 (continued)

Question	Response		Secondary School Appraisers (N=64)	University Appraisers (N=18)
	Yes	No		
4. Four criteria will be used to judge the composition:				
(i) Total number of clauses				
(ii) Grammar				
(iii) Extent of vocabulary				
(iv) Accuracy in the use of vocabulary (not including spelling or use of accents)				
Are there other criteria of equal or greater importance?				
	Yes	No	20	7
			35	7

Note: Omitted responses account for discrepancies between the number of tabulated responses and the number of appraisers.

TABLE A3.18

A Tabulation of Responses to a General Question About
Use of Tests Like the Speaking Test

<u>Question</u>	<u>Responses</u>		Secondary School Appraisers (N = 64)	University Appraisers (N = 18)
Considering the <u>Speaking Test</u> as a whole, do you use this kind of test to evaluate the ability of your students to speak French?	Yes, frequently		9	8 ^a
	Yes, occasionally		33	
	No		15	6

Note: Omitted responses account for discrepancies between the number of tabulated responses and the number of appraisers.

^aThe appraisal inventory for university instructors offered only one unqualified response of Yes.

TABLE A3.19

Percentage of Students Responding Correctly to Each Item in the
Reading Test and the Listening Test

<u>Item No.</u>	<u>Reading Test</u>	<u>Listening Test</u>
1	97	56
2	81	80
3	62	85
4	78	86
5	69	89
6	41	24
7	83	52
8	49	87
9	31	86
10	32	81
11	63	52
12	78	53
13	60	54
14	38	79
15	45	61
16	13	52
17	26	91
18	95	71
19	88	82
20	82	36
21	97	68
22	85	58
23	85	56
24	90	43
25	81	61
26	54	67
27	73	59
28	89	63
29	87	51
30	86	63
31	48	49
32	62	87
33	24	70
34	30	79
35	66	
36	85	
37	78	
38	32	
39	52	

TABLE A3.20

Frequency Distributions of Difficulty Indices for Items
in the Reading and Listening Tests

<u>Interval for Index</u>	<u>Reading Test</u>	<u>Listening Test</u>
90-99	4	4
80-89	11	9
70-79	4	4
60-69	6	7
50-59	2	9
40-49	4	2
30-39	5	1
20-29	2	1
10-19	1	
Number of items	39	34
Mean	64	66
S.D.	24	16

TABLE A3.21

Percentage of Students Omitting and Not Reaching Each Item in the
Reading Test and Omitting Each Item in the Listening Test

<u>Item No.</u>	<u>Reading Test</u>		<u>Listening Test</u>	
	<u>% Omitted</u>	<u>Not Reached</u>	<u>% Not Reached</u>	<u>Omitted</u>
1	1		0	0
2	3		0	0
3	12		0	3
4	7		0	3
5	1		0	1
6	8		0	8
7	7		0	1
8	11		0	0
9	19		0	1
10	37		0	9
11	8		0	5
12	8		0	10
13	19		0	8
14	36		0	7
15	4		0	7
16	17		0	16
17	31		0	0
a-----				
18	0		0	2
19	2		0	0
20	3		0	41

TABLE A3.21 (continued)

<u>Item No.</u>	<u>Reading Test</u>		<u>Listening Test</u>	
	<u>% Omitted</u>	<u>% Not Reached</u>	<u>% Omitted</u>	
21	0	0	4	
22	3	0	6	
23	0	0	0	
24	2	0	26	
25	1	0	7	
26	8	0	9	
27	10	0	9	
28	1	0	2	
29	2	0	9	
30	2	0	1	
31	11	1	4	
32	10	1	0	
33	25	2	6	
34	25	3	1	
35	2	5		
36	3	6		
37	0	7		
38	0	19		
39	0	19		

^aThis is the point separating the items in the first part of the Reading Test from the items in the second part of the test.

TABLE A3.22

Means, Standard Deviations and Ranges of Marks Assigned by Each
of Two Markers to 54 Writing Test Papers

<u>Variable</u>	<u>Marker</u>	<u>Mean</u>	<u>S.D.</u>	<u>Low</u>	<u>High</u>
Total Writing I	1	20.81	7.54	6.00	38.00
	2	20.78	7.49	7.00	38.00
Number of Clauses	1	13.89	5.23	2.00	28.00
	2	14.30	6.10	1.00	26.00
Number of Grammatically Correct Clauses	1	6.57	4.60	1.00	21.00
	2	6.91	4.62	0.00	20.00
Number of Clauses with Correct Vocabulary	1	10.48	4.60	1.00	27.00
	2	11.09	5.27	0.00	25.00

TABLE A3.23

Correlation Coefficients Between the Scores Assigned
by Each of Two Markers to 54 Writing Test Papers

<u>Variable</u>	<u>Coefficient</u>
Total Writing I	0.99
Number of Clauses	0.93
Number of Grammatically Correct Clauses	0.84
Number of Clauses with Correct Vocabulary	0.84

TABLE A3.24
Means and Standard Deviations of the Distributions of
Different Writing Test Scores
(N = 54)

<u>Variable</u>	<u>Mean</u>	<u>S.D.</u>
Part I:		
Total Writing I	20.8	7.5
Part II:		
Number of Clauses	14.1	5.6
Number of Grammatically Correct Clauses	6.7	4.4
Number of Clauses with Correct Vocabulary	10.8	4.7
Proportion of Grammatically Correct Clauses	0.45	0.18
Proportion of Clauses with Correct Vocabulary	0.75	0.13
Writing Quantity	41.8	21.6
Writing Quality	40.5	9.1
Total Writing II	82.3	29.1

TABLE A3.25

Coefficients of Intercorrelation Among the Different Scores on the
Writing Test
(N = 54)

Variable	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>
Part I:								
1. Total Writing I								
Part II:								
2. Number of Clauses	52							
3. Number of Grammatically Correct Clauses	65	80						
4. Number of Clauses with Correct Vocabulary	55	93	80					
5. Proportion of Grammatically Correct Clauses	57	44	85	49				
6. Proportion of Clauses with Correct Vocabulary	30	22	27	52	24			
7. Writing Quantity	64	90	96	93	73	39		
8. Writing Quality	59	45	80	62	91	63	76	
9. Total Writing II	66	81	96	88	83	49	98	88

Note: Decimal points have been omitted.

TABLE A3.26

Means, Standard Deviations and Minimum and Maximum Scores of the
Distribution of Speaking Test Scores
(N=50)

<u>Variable</u>	<u>Mean</u>	<u>S.D.</u>	<u>Min-Max</u>
Pronunciation	21.0	3.4	13- 29
Structural Control	19.3	8.4	0- 36
Oral Reading	15.9	3.7	9- 23
Fluency:			
Global Rating	1.5	1.1	0- 4
Number of Clauses	12.8	4.1	5- 24
Number of Different Grammatical Structures	8.6	3.4	3- 18
Number of Clauses Correct in Structure	9.8	4.2	1- 21
Number of Clauses Correct in Morphology	7.3	4.6	0- 19
Number of Clauses Correct in Vocabulary	9.3	4.4	1- 21
Number of Clauses Correct in Pronunciation	9.5	4.7	1- 22
Quantity	81.7	35.3	18-182
Quality	14.7	3.3	5- 20
Total	96.3	37.8	23-200
<u>Total Speaking Test</u>	208.8	57.9	97-346

TABLE A3.27

Reliability of Fluency Scores on the Speaking Test

<u>Variable</u>	<u>Half-Test Correlation</u>	<u>Reliability^a</u>
Number of Clauses	.61	.76
Number of Different Grammatical Structures	.61	.76
Number of Clauses Correct in Structure	.58	.73
Number of Clauses Correct in Morphology	.63	.77
Number of Clauses Correct in Vocabulary	.61	.76
Number of Clauses Correct in Pronunciation	.63	.77

^aEstimated from the half-test correlation by the Spearman-Brown formula
(Lord and Novick, 1968, p. 84)

TABLE A3.28
Intercorrelation of Scores on the Speaking Test
(N = 50)

<u>Variable</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>	<u>12</u>	<u>13</u>
1. Pronunciation													
2. Structural Control	54												
3. Oral Reading	57	52											
Fluency Part of Test													
4. Global Rating	38	68	64										
5. Number of Clauses	25	32	41	58									
6. Number of Different Grammatical Structures	30	43	51	71	80								
7. Number of Clauses Correct in Structure	27	39	49	69	93	80							
8. Number of Clauses Correct in Morphology	35	56	53	77	79	77	81						
9. Number of Clauses Correct in Vocabulary	34	56	49	71	89	77	89	84					

TABLE A3.28 (continued)

<u>Variable</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>	<u>12</u>	<u>13</u>
10. Number of Clauses Correct in Pronunciation	43	55	58	79	84	79	85	79	86				
11. Quantity	37	54	56	79	92	89	95	89	95	94			
12. Quality	35	67	55	78	46	56	65	71	71	75	73		
13. Total	37	56	57	80	90	88	94	90	95	94	99	77	
14. Total <u>Speaking Test</u>	59	79	72	85	76	80	82	86	88	90	92	81	93

Note: Decimal points have been omitted.

TABLE A3.29

Discrimination Indices for Items in the
Reading Test and the Listening Test

Item	Discrimination Index ^a		Discrimination Index ^a	
	Reading Test (N=653)	Listening Test (N=534)	Item	Reading Test Listening Test
1	40	41	21	21 70
2	70	47	22	47 48
3	72	61	23	48 58
4	58	54	24	53 56
5	57	53	25	66 70
6	52	62	26	73 69
7	41	61	27	60 55
8	49	64	28	59 72
9	49	77	29	67 70
10	55	75	30	63 73
11	60	72	31	50 48
12	46	57	32	58 65
13	65	71	33	61 54
14	69	68	34	58 69
15	54	67	35	55 -
16	37	71	36	48 -
17	42	40	37	51 -
18	27	72	38	67 -
19	41	33	39	38 -
20	53	51	Mean	61
			S.D.	11

Note: Decimal points have been omitted.

^aThe index that is reported is the biserial correlation between item scores and total test scores, uncorrected for the inclusion of the item in the test.

TABLE A3.30

Distribution Statistics and Information on the Reliability
and Standard Error of Measurement of the
Reading Test and the Listening Test

	<u>Reading Test</u>	<u>Listening Test</u>
Number of Examinees	653	534
Number of Test Items	39	34
Mean ^a	22.66	20.29
S.D. ^a	7.70	8.51
Highest Score	39.00	34.00
Lowest Score ^b	-2.33	-4.67
Reliability ^c	0.85	0.88
Standard Error of Measurement ^c	2.99	2.92

^aThis statistic is for the distribution of corrected for guessing scores before any adjustment was made for the effect of absenteeism.

^bNegative scores arise because a correction for guessing was applied; wrong answers were scored -1/3 and correct answers were scored 1. Omitted questions were ignored in scoring.

^cEstimated from a formula due to Hoyt (1941).

APPENDIX A4

TECHNICAL REPORT ON THE TEST

DE CONNAISSANCE DE LA LANGUE (ANGLAIS)

This test was prepared for administration to Francophone students at the SSGD and SSHGD levels to assess selected aspects of their competence in the use of English. The purpose of this report is to describe the contents of the test in some detail, to summarize the responses of a group of secondary and postsecondary instructors who appraised the test, and to provide some information about technical matters related to the test.

1. DESCRIPTION OF TEST CONTENT

The Test de connaissance de la langue (anglais) contained two parts. The first part assessed some aspects of the ability to read English for comprehension. The second part was designed to assess some aspects of the ability to write in English.

1.1 Reading Comprehension Part

This part of the test, which was prepared at the University of Michigan, consisted of four passages, each accompanied by five multiple-choice questions. The passages ranged in length from approximately 150 words to approximately 200 words and dealt with the following topics: Viking invaders of England, systems of writing, describing a sailor, and poetry and the arts. The

multiple-choice questions associated with the passages could be classified roughly as assessing three different types of comprehension: literal understanding of a passage; identification of the main purpose of a passage, and ability to draw inferences or conclusions from a passage.

1.2 Writing Exercise

This part of the test consisted of a passage written in English. The student was to read the passage, and then to write a 100-to 150-word summary of the main points in the passage. Finally, the student was to write a brief statement of his own opinion about the issue dealt with in the passage.

Obviously, more than writing ability, however defined, was required to perform the "writing" test. The added component was, again, ability to read for comprehension. The passage which formed the basis of the test was part of an article published in a Canadian magazine. At a length of approximately 550 words, it was considerably longer than the four passages contained in the "reading comprehension" part of the test.

2. TEST APPRAISAL

The Test de connaissance de la langue (anglais) and the related Test Appraisal Inventory were sent to practising teachers as a means of sampling professional opinion about the test. Among those asked to respond were teachers in the fourteen Francophone secondary schools of the study who taught English or anglais courses to Grade Twelve and Grade Thirteen students. In addition, the test and inventory were sent to some of the instructors in two Ontario universities and three Colleges of Applied Arts and Technology teaching English to Francophone students in first year university and college courses. Because there was no attempt to draw a probability sample of instructors at either the secondary

or the postsecondary level, the responses that were received cannot be used to draw inferences about the opinions held by the population of individuals in the province giving English language instruction to Francophones. The responses can only be said to constitute the collective opinions of the 41 secondary school teachers and the 18 college and university instructors who undertook to comment on the tests.

A tabulation of coded responses to each question contained in the inventory is presented in Table A4.1. There are several features of the table that deserve mention here: (i) The secondary school teachers responded in terms of courses for Francophones in English or anglais at either the Grade Twelve or Grade Thirteen level. Thus, the responses of the 41 secondary school teachers are divided among four columns, one column for each type of course at each grade level. The number of responses to a question when summed over the four columns, totals more than 41, however, because some teachers responded for two or more courses of different types and grade levels. (ii) Secondary courses were not differentiated by "general" and "advanced" designation. Although this would have been desirable, it would have made the inventory more complex and time-consuming to complete. (iii) The responses of college and university instructors were not separated for the purpose of tabulating results but appear in a single column of the table. Too few instructors from either colleges or universities responded to make such a separation worthwhile in the sense of revealing meaningful differences. (iv) The postsecondary appraisal inventory overlapped the secondary inventory but contained fewer questions, a fact that accounts for the absence of postsecondary tabulations for questions 15, 16, 17, 22 and 23. (v) The number of responses that were tabulated is not necessarily the same from question to question for either the secondary or the postsecondary group because some questions were omitted by some respondents.

In addition to coded responses, the instructors who completed the appraisal were invited to offer written comments at appropriate places throughout the questionnaire. These comments

have been used to elaborate the impressions provided by the coded responses.

2.1 Reading Comprehension Part

Responses by instructors at both the secondary and postsecondary levels to the first question in the inventory (see Table A4.1) indicated good support for the inclusion of reading comprehension in a test of English for Francophones. Regarding the four reading passages included in the test, there was general agreement among all instructors, regardless of level of course being considered, that the difficulty of the reading passages as a set was about right (see tabulation for Question 2, Table A4.1). When judged on an individual basis, the first reading passage was regarded as somewhat easy, whereas the second and fourth passages were thought to be somewhat difficult, (see tabulations for questions 3,4 and 6). Despite this, the majority of instructors agreed that the four passages were representative of the material students in Grades Twelve and Thirteen should be able to read with comprehension. Moreover, the tabulation of responses to Questions 8-11 indicated that the majority of instructors, again without regard for level of course, considered none of the passages as inappropriate for reasons other than difficulty. Written comments from those instructors who did object to a passage for reasons other than difficulty covered a variety of complaints. These included an objection to the first passage because it was "culturally loaded" and therefore uninteresting; an observation that the second passage involved a technical subject with specialized vocabulary; and a comment that the fourth passage was a "foolish piece, filled with simplifications and popularizations". It must be remembered, however, that these were reactions from a minority of the appraisers. In fact, one appraiser said of a passage that it was similar to what would be expected in literature and composition classes at the postsecondary level.

The multiple-choice questions about the reading passages could be characterized as assessing (i) the ability to comprehend the literal meaning of a passage, (ii) the ability to identify the main idea or objective of a passage, and (iii) the ability to draw conclusions or implications from a passage. Instructors at all levels were asked to estimate the number of students who--on entry into their courses--could read successfully for any one of these purposes. The tabulations for Questions 12, 13 and 14 of Table A4.1 support the reasonable expectation that, as the level of instruction increases from Grade Twelve through Grade Thirteen to postsecondary, the level of expectation of instructors for what students can do when they enter a course also increases, and that expectations of instructors, regardless of level of instruction, are higher for literal comprehension ability than for ability to identify the main idea or objective; expectations for the latter ability are, in turn, higher than for the ability to draw conclusions and implications. When these results are coupled with the tabulations of responses that secondary instructors gave to questions about the degree of attention accorded reading for literal comprehension, for identification of main ideas, and for conclusions and implications (see the tabulations for Questions 15, 16 and 17), it becomes clear that, however limited the set of reading comprehension abilities assessed may have been, they are the skills that the group of secondary instructors who responded emphasize in their courses and that the students of the postsecondary instructors who responded are expected to possess in some measure on entry into first year courses.

The limitations of the reading comprehension test were emphasized in the responses to Questions 18 and 19. On the order of one-half the instructors, regardless of level, indicated that there were important aspects of reading comprehension not measured by this test. In their written responses, they indicated that the test did not assess such things as a student's understanding of vocabulary and his reading speed. One respondent even questioned whether or not the test items did actually assess ability to interpret or draw inferences, as had been suggested in the test appraisal inventory. There was a strong body of opinion among

these instructors that multiple-choice tests of the sort used to assess reading comprehension were inefficient and inadequate measures of the three abilities noted above. Concern was expressed about the effects of chance success in the scores obtained on the tests, and about the need for some kind of written exercise to assess understanding and ability to interpret.

If there are general conclusions to be drawn about the reading comprehension part of the Test de connaissance de la langue (anglais), they would seem to be (i) that the test was limited in obvious ways because of the use of multiple-choice questions, and (ii) that specific aspects of certain passages and questions were bothersome to one or another of the instructors who appraised the tests, but (iii) that, when these limitations are taken into account, the test was a satisfactory means of assessing some important components of reading comprehension ability.

2.2 Writing Exercise

The Test de connaissance de la langue (anglais) posed, in a second part, the task of reading a passage, summarizing its contents in a précis of 100-150 words, and writing a commentary of 100-150 words on the points made by the author of the passage. As the tabulation of responses to Question 20 reveals, most instructors felt that a composition task is either important or very important in an assessment of language competence. The tabulation of responses to Question 21 indicates that the composition task included in the test was one that most instructors would expect most students to be able to do when they entered a particular course, although expectations were higher for students entering Grade Thirteen and postsecondary courses than for students entering Grade Twelve courses. The responses of secondary instructors to Question 22 suggest that development of the skills needed to perform a composition exercise such as the one that was included in the test is an objective of virtually all the

secondary instructors, although degree of emphasis varies to a considerable extent.

A further indication of the suitability of the writing exercise is provided by responses to Questions 24 and 25. The majority of instructors agreed, in response to Question 24, that the exercise was pitched at an appropriate level of difficulty, although the extent of agreement was, not unexpectedly, somewhat less in the responses for Grade Twelve courses. In response to Question 25, most instructors agreed that a single writing exercise was not an unfair basis for assessing students, although there were written comments in which the particular writing exercise used in the test was criticized. Among the written comments were the following suggestions:

- (a) that the student be offered a choice of topic--choice being essential when personal opinion--even on general interest subjects--is asked for;
- (b) that the writing exercise be supplemented with an oral test so that questions could be asked about what a student had written;
- (c) that the writing test be supplemented with a listening test in which the student took notes, much as in lecture classes; and
- (d) that a speaking exercise be added to the test.

Question 25 of the appraisal inventory asked the instructors to rate various criteria for grading the writing exercise. The tabulation of ratings provides evidence of the extent of disagreement amongst the instructors as to the most important quality of good writing--structure; logical presentation; style; or grammar, usage and mechanics. Only the criteria "choice-of-words" was rated low by most secondary instructors and relatively low by postsecondary instructors. In their written comments, the respondents drew attention to such things as unity,

coherence, and clarity as other criteria of importance. Another respondent commented that it was difficult for him to distinguish criteria when it came to judging a student's composition and that the most important thing was "clean, succinct communication".

Other written responses about the writing exercise pointed to its difficulty and to the limitation of the length of the précis to only 100-150 words. A few instructors thought that the exercise was too difficult and that a précis of 200 words should have been judged acceptable.

In conclusion, it seems fair to say that the respondents to the Test Appraisal Inventory agreed on the need for a writing exercise in a test of language competence and felt that the writing exercise used in the test was an appropriate one for assessing some important aspects of language competence.

2.3 Multiple-Choice Questions vs Writing Exercise

Four questions at the end of the appraisal inventory solicited opinions about the use of multiple-choice tests for assessing various aspects of language competence and about the need for both a writing exercise and a multiple-choice test. The conclusion that may be framed in light of the tabulations of responses to Questions 27, 28, 29 and 30 is that the respondents saw a place for multiple-choice questions but that they thought that, when language competence was being assessed by means of both multiple-choice questions and a writing exercise, the assessment provided by the writing exercise should be accorded more weight.

3. TECHNICAL ISSUES

3.1 Scoring the Reading Comprehension Part

This part of the test contained 20 multiple-choice items, each having four options. As indicated in the instructions to students taking this test, a correction for guessing was applied. This correction was the standard one for four-choice items, giving each incorrect answer a weight of $-1/3$. A correct response was scored 1 and omitted questions were ignored in the scoring.

3.2 Scoring the Writing Exercise

The writing exercise consisted of a passage to be summarized and commented upon. This exercise was scored for the adequacy of the summary and for the quality of the commentary. In scoring the summary, markers awarded one mark for each idea in the passage, up to a maximum of 10, and deducted marks for excessive length and for misrepresentation of ideas in the passage. The commentary was scored holistically on a scale from 1 for low quality to 10 for high quality.

The task of scoring the writing exercises was accomplished as follows. All exercises except 20 were scored by three different markers. The special set of 20 exercises was scored by all six of the persons engaged to do the marking. Scores on the special set of 20 exercises were used to implement a procedure for adjusting the scores assigned by different markers. Adjustments were necessary to remove differences among the markers in the average mark assigned, the range of marks assigned and the reliability of marking. The procedure used to adjust marks is described in Appendix D2. In the end each writing exercise had assigned to it only two scores, one for the summary and one for the commentary, each representing a synthesis of the marks assigned by the different markers.

3.3 Item Difficulty in the Reading Comprehension Part

Difficulty is indexed as the average over the Francophone schools in the study of the percentage of students within a school who responded correctly to a multiple-choice item. Given the sampling design of the study, this percentage constitutes an estimate of the fraction of students in the Francophone population who would be able to answer the question correctly. (For a definition of this population, see Chapter Two of this report.)

The difficulty index of each item for Grade Twelve students and for Grade Thirteen students is reported in Table A4.2. A mean and standard deviation of the difficulty indices is also reported for each grade. It is apparent from these statistics that the reading comprehension part of the Test de connaissance de la langue (anglais) was relatively easy for students in either grade, although it was clearly easier for the average Grade Thirteen student than for the average Grade Twelve student. This conclusion is reinforced by the frequency distribution of item difficulty indices that appear in Table A4.3. All items except one have percentages of correct responses above 50 for the Grade Twelve students and above 60 for the Grade Thirteen students. In conclusion, if the objective of testing is to spread students out just as much as possible in terms of their performance on the test, then the reading comprehension part of the Test de connaissance de la langue (anglais) was too easy for Grade Thirteen students. An average item difficulty somewhat higher than 0.60, but not as much higher as 0.76, can be regarded as ideal.

3.4 Speededness of the Reading Comprehension Part

The time allowed students to work on the reading comprehension part was 30 minutes. This appears to have been more than enough time for the vast majority of students to attempt all the items in the test. In analyses of multiple-choice tests, the convention sometimes followed is arbitrarily to decide that a student has

not reached an item if he fails to answer it and all succeeding items in the test. Otherwise, an item that has not been answered is classified as an omitted item. In the analysis of Grade Twelve responses to the reading comprehension part of the Test de connaissance de la langue (anglais), no omission of an item was classified as not reached, except for omissions of the very last item. It is a matter of choice whether to class failures to respond to the last item in a test as evidence of too little working time or as evidence of an inability to identify the correct answers. In the analysis of Grade Thirteen responses, only one per cent of students could be classified as not having reached the second last item and six per cent were classified as not having reached the last item.

Additional evidence on the lack of speeding is contained in the percentages of omissions that are reported in Table A4.2. For every item, these are very small or zero, suggesting that students had ample time to go back over their work and attempt questions that were not answered the first time through the test.

In general, then, it can be concluded that speededness was not a problem on the reading comprehension part of the Test de connaissance de la langue (anglais).

3.5 Reading Comprehension Part: Information on Discrimination, Reliability and Distribution of Scores

Discrimination can be loosely defined as the extent to which performance on an item is correlated with overall performance on a test. In tests designed to spread students over the range of possible scores, it is generally held that items should have high indices of discrimination. The biserial coefficient of correlation between scores on an item and scores on the test of which the item is a part is used as an index of discrimination. One rule of thumb for assessing this index is to judge it in relation to 0.3: indices higher than this are acceptable, indices lower than this are unacceptable.

The indices of discrimination for items in the reading comprehension part of the Test de connaissance de la langue (anglais) are reported in Table A4.4. These figures were derived from an analysis of the responses of all the students at both the SSGD and SSHGD levels who wrote the test. It can be seen that the indices reported in Table A4.4 are all above 0.3. A conclusion regarding the acceptability of the item discrimination indices for this test is obvious.

Other information about the reading comprehension part of the Test de connaissance de la langue (anglais) is provided in Table A4.5. The fact that this part of the test was relatively easy for the group of students who wrote it is reflected in the relatively high mean. As might be expected, given that the test was easy, the distribution of scores (not shown) was negatively skewed. One indication of this skewness was the fact that 15 students answered all 19 questions correctly, but only three students had below-zero scores. (Negative scores reflect a level of performance below that which would be expected if responses were made entirely at random.)

As regards reliability, the figure 0.71 is adequate given the brevity of the test and the purposes for which it was intended in this study. These purposes were to make comparisons among different groups of students in terms of their performance on the test and to predict school marks.

3.6 Reliability of Scoring the Writing Exercise

An estimate of the reliability with which the writing exercise of the Test de connaissance de la langue (anglais) was scored was obtained from a factor analysis involving the 20 exercises that were scored by all six markers. Using scores on these exercises as observations on the markers, a 6 by 6 matrix of the coefficients of intercorrelation among the markers was obtained. When this matrix was factor analyzed, a single common factor was obtained. The squared factor loading of a marker on this factor

constitutes a lower-bound estimate of the reliability with which the marker scored the essays. (The theoretical rationale underlying this assertion is provided in Harman, 1960, Chapter 2.)

Two estimates of reliability were obtained in this way for each marker: the first was an estimate of the reliability with which the summary portion of the writing exercise was scored; the second estimate was of the reliability with which the commentary part of the exercise was scored. These estimates of reliability are reported in Table A4.6.

It is apparent from the numbers presented in Table A4.6 that the summary part of the writing exercise was scored much less consistently than the commentary part. This may be a reflection of the different scoring procedures that were followed for each part. The summary was scored by counting the number of ideas in the summary that were present in the paragraph that was to be summarized, and by making deductions for excessive length and misrepresentation of ideas. The commentary, on the other hand, was to be scored holistically, by assigning scores on a scale from 1 to 10 so as to represent the quality of the commentary. Obviously, the markers achieved much higher agreement in scoring the commentary, using the apparently more subjective, holistic judgment approach, than in scoring the summary, using the seemingly more objective approach of counting ideas.

The question arises of whether these estimates of reliability are adequate for the purposes of the study. These purposes consist of making comparisons among various subgroups of students and predicting school marks. It is obvious that the scores on the commentary, with their higher reliability of scoring, will serve these purposes better than the scores on the summary. The relatively low reliability of the summary scores means that the "confidence interval" for an intergroup difference will be larger than it would have been had a higher degree of reliability been achieved. Also, the variance of errors of estimate about the line regressing school marks on test scores

will be larger than it would have been had the reliability been higher.

The less reliable the measurement of a variable, the more that chance (random) factors affect the results and the less likely it is that observed differences between groups or observed coefficients of regression will be judged to be significant, in a statistical sense. Nevertheless, the relatively low reliability of the summary scores does not provide grounds for excluding them from the study; they may be found to yield results of interest despite their low reliability.

REFERENCES

Harman, H.H. Modern factor analysis. Chicago: University of Chicago Press, 1960.

Hoyt, C. Test reliability estimated by analysis of variance. Psychometrika, 1941, 6, pp. 153-160

TABLE A4.1

FREQUENCY TABULATION OF CODED RESPONSES IN APPRAISALS OF THE
TEST DE CONNAISSANCE DE LA LANGUE (ANGLAIS)

		Secondary (N=41)				Post- Sec. (N=18)
		Grade 13		Grade 12		
		English	Anglais	English	Anglais	
1.	Do you consider that a test of reading comprehension, not necessarily the one used, assesses an important component of language achievement at this level or at the beginning of post-secondary study?					
	(1) YES.....	12	5	23	10	18
	(2) NO.....	3	1		1	
<u>Questions 2-6:</u> Apply the following responses and corresponding response codes:						
	(1) too easy					
	(2) somewhat easy					
	(3) about right					
	(4) somewhat difficult					
	(5) too difficult					
		Response Codes				
2.	What is your overall	1	1	2		
	assessment of the	2	3	2	2	4
	difficulty level of four	3	10	3	14	7
	reading comprehension	4		4	3	5
	passages?	5		1	1	1

TABLE A4.1 (Continued)

		Response Codes (see previous page)	Secondary				Post- Sec.
			Grade 13		Grade 12		
			English	Anglais	English	Anglais	
Assess the difficulty of each passage separately in terms of the above set of responses:							
3. Passage on Viking invaders	1	2		2		3	
	2	9	4	9	2	4	
	3	2	1	10	6	7	
	4			2	2	2	
	5					2	
4. Passage on systems of writing	1			1		1	
	2	4	3	4	2	1	
	3	4		6	1	4	
	4	5	2	8	5	10	
	5			4	2	2	
5. Passage describing the sailor	1	1				1	
	2	5	4	6	1	4	
	3	7	1	8	8	9	
	4			8	1	3	
	5			1		1	
6. Passage on poetry and the arts	1					1	
	2	3	1	3	1		
	3	4	3	8	2	10	
	4	6	1	7	6	6	
	5			5	1	1	
7. Are the four passages representative of the material you would expect students at this level to be able to read with comprehension?							
(1) YES.....		10	5	14	4	12	
(2) NO.....		3		8	5	6	

TABLE A4.1 (Continued)

	Secondary				Post- Sec.
	Grade 13		Grade 12		
	English	Anglais	English	Anglais	
<u>Questions 8-11: Is any passage inappropriate for reasons other than difficulty?</u>					
8. Passage on Viking invaders					
(1) YES.....	3		1	1	6
(2) NO.....	10	5	21	8	12
9. Passage on systems of writing					
(1) YES.....	2		8	3	3
(2) NO.....	11	5	14	6	15
10. Passage describing the sailor					
(1) YES.....			1	2	3
(2) NO.....	13	5	21	7	15
11. Passage on poetry and the arts					
(1) YES.....	2	1	5	2	6
(2) NO.....	11	4	17	7	12
Generally the items appear to test the student's ability in the following areas:					
I. Literal understanding of the passage					
II. Identification of main idea or purpose					
III. Inference or implication					

TABLE A4.1 (Continued)

Questions 12-14 concern the issue of whether students should have these three abilities on entry to English or Anglais courses at these levels or before entry to post-secondary study. Select your answers to these questions from among the following options:

(1) All the students

(2) More than 75% of the students, but not all

(3) 51% to 75% of students

(4) 26% to 50% of students

(5) One or more, but less than 26%

(6) None

12. How many of the students entering courses at this level should have the ability to read a passage for literal understanding?

13. How many students should be able to identify the main idea or purpose of a passage?

14. How many students should be able to draw inferences and see implications?

Response Codes	Secondary				Post-Sec.
	Grade 13		Grade 12		
	English	Anglais	English	Anglais	
1	7	1	5		16
2	7	3	11	4	
3		1	6	4	
4			2	1	
5				1	
6					
1		1	6		12
2	11	3	7	3	3
3	2	1	7	5	
4	1		3	2	1
5					
6			1		
1		1	2		8
2	4	3	6	2	3
3	9	1	6	4	2
4			5	4	1
5			4		1
6					

TABLE A4.1 - (Continued)

		Secondary				Post-Sec.
		Grade 13		Grade 12		
		English	Anglais	English	Anglais	
<p>Questions 15-17 ask for an assessment of the amount of emphasis each of the three reading abilities receives in the courses of each type that you teach. Select your answer to these questions from among the following options:</p> <p>(1) Heavy emphasis</p> <p>(2) Moderately heavy emphasis</p> <p>(3) Light emphasis</p> <p>(4) Individual, remedial emphasis only</p> <p>(5) No emphasis</p>						
	Response Codes					
15.	What emphasis is given to reading for <u>literal understanding</u> ?	1 2 3 4 5	3 5 3 3	1 3 1	10 6 2 5	2 3 5
16.	What emphasis is given to reading <u>to identify the main idea or purpose</u> ?	1 2 3 4 5	7 7	1 4	9 8 6	5 4 1
17.	What emphasis is given to <u>drawing inferences and seeing implications</u> ?	1 2 3 4 5	4 9 1	3 2	9 5 8 1	3 4 3

TABLE A4.1 (Continued)

WRITING EXERCISE

Note: In responding to questions 20-23 inclusive, do not refer in particular to the writing task assigned in this study.

Secondary				Post-Sec.
Grade 13		Grade 12		
English	Anglais	English	Anglais	
5	2	7	2	14
8	2	14	5	3
	1	2	3	1
2		3		7
6		3		8
5	5	11	5	1
		2	4	
		3		
		1	1	

TABLE A4.1 (Continued)

Secondary					Post-Sec.
Grade 13		Grade 12			
	English	Anglais	English	Anglais	
18. Are there important reading skills that have not been tested in the tests used but that should have been?					
(1) YES.....	6		9	2	9
(2) NO.....	5	5	9	6	8
19. Is the multiple-choice format a reasonable method of assessing at least those three reading comprehension abilities mentioned above?					
(1) YES.....	7	2	12	2	8
(2) YES (Qualified).....	5	3	8	7	8
(3) NO.....	2		3		1

TABLE A4.1 (Continued)

		Secondary				Post-Sec.
		Grade 13		Grade 12		
		English	Anglais	English	Anglais	
22.	How much emphasis do you give in your teaching to the development of student competence in this type of writing?					
	(1) Heavy emphasis.....	3		5	2	
	(2) Moderately heavy emphasis.....	8	2	12	3	
	(3) Light emphasis.....	2	3	6	4	
	(4) No emphasis.....				1	
23.	How many of the students who successfully complete courses at this level should be able to produce an acceptable piece of writing of this type?					
	(1) All.....	4		3		
	(2) More than 75%, but not all.....	8	3	9	3	
	(3) 51% to 75%.....	1	2	5	3	
	(4) 26% to 50%.....			4	3	
	(5) One or more, but less than 26%.....			1		
	(6) None.....			1	1	

TABLE A4.1 (Continued)

Note: Answer questions 24-26 inclusive with reference to the specific writing assignment given.

24. Was the assignment at a reasonable level of difficulty for students in courses at this level or upon entry to college or university?

(1) YES.....
(2) NO.....

25. Was the restriction to a single topic unfair to students in a test of writing competence?

(1) YES.....
(2) NO.....

26. Rank, in order of importance, the following criteria for evaluating this type of writing. Using the numbers 1, 2, 3, 4, 5, (1-High, 5-Low) write the ranks in the five spaces provided in each column.

1. Organization

Response Codes

1
2
3
4
5

2. Logic, use of evidence

1
2
3
4
5

Secondary				Post-Sec.
Grade 13		Grade 12		
English	Anglais	English	Anglais	
13	3	17	4	16
	2	6	4	2
2	2	6	5	2
11	3	17	4	16
5	4	11	5	6
1		1	3	7
2	1	4	1	1
3	1	1	1	
3		3		4
5	1	6	1	6
5	3	9	4	5
3		4	2	2
1	1	2		5
	1		3	

TABLE A4.1 (Continued)

		Secondary				Post-Sec.
		Grade 13		Grade 12		
		English	Anglais	English	Anglais	
3. Style (chiefly the sentence)	1		1	1	1	4
	2	5	1	5	2	2
	3	6	1	7	2	4
	4	3	1	3	4	1
	5		2	4	1	7
4. Grammar, usage, mechanics	1	3	2	2	3	8
	2	2	1	6	3	2
	3	2	3	5	3	4
	4	6		7	1	4
	5	1				
5. Diction	1	1	1	1		1
	2	1	2	1	2	3
	3	1	1	3	2	5
	4	1	1	5	3	5
	5	10	1	10	3	4

TABLE A4.1 (Continued)

OTHER QUESTIONS	Secondary				Post- Sec.
	Grade 13		Grade 12		
	English	Anglais	English	Anglais	
27. Do you think a test in multiple-choice format is in general a satisfactory means of assessing students' competence in such areas as style, grammar, sentence structure and vocabulary?					
(1) YES.....	3	1	3		3
(2) YES (Qualified).....	2	3	5	5	7
(3) NO.....	8	1	13	5	6
28. Where multiple-choice tests are used, should they be supplemented by other measures?					
(1) YES.....	11	4	18	10	16
(2) NO.....	2	1	2		2
29. How do you regard the use of both a multiple-choice test of language achievement and a sample of writing in assessing language competence?					
(1) The multiple-choice test is satisfactory by itself.....	1				
(2) The use of both is important.....	9	4	18	8	15
(3) The writing sample is satisfactory by itself.....	1	1	1	1	2
(4) Neither is particularly satisfactory.....				1	1

TABLE A4.1 (Continued)

30. If a student's relative standing on a writing test were different from the student's relative standing on a multiple-choice test of language achievement, which would you consider the most valid measure of the student's language competence?

- (1) the score on the writing test.....
- (2) the score on a multiple-choice test.....
- (3) a combined score weighted in favour of the writing test.....
- (4) a combined score weighted in favour of the multiple-choice test.....
- (5) a combined score giving equal weight to both tests.....

Secondary				Post-Sec.
Grade 13		Grade 12		
English	Anglais	English	Anglais	
2	1	3	2	3
7	3	12	5	11
1			2	2
3	1	6	1	2

TABLE A4.2

Percentage of Students at Each of Two Grade Levels
Who Correctly Answered and Who Omitted Questions in
the Reading Comprehension Part of the Test de
connaissance de la langue (anglais)

Item No.	Grade 12		Grade 13	
	<u>Percent Correct</u>	<u>Percent Omitting</u>	<u>Percent Correct</u>	<u>Percent Omitting</u>
1	82	2	90	0
2	56	0	61	3
3	55	3	74	4
4	78	2	84	2
5	81	0	93	0
6	63	0	83	0
7	25	8	28	5
8	69	1	80	2
9 ^a	-	-	-	-
10	51	2	65	1
11	85	0	89	1
12	55	0	74	1
13	97	0	98	0
14	65	1	75	1
15	57	1	67	1
16	74	0	80	0
17	55	2	66	3
18	59	1	70	1
19	87	0	94	1
20	61	6	78	6
Mean	66		76	
S.D.	16		15	

^aResults are not reported for this item inasmuch as it was judged to have two correct answers.

TABLE A4.3

Frequency Distributions of Difficulty Indices
for Items in the Reading Comprehension Part of the
Test de connaissance de la langue (anglais)

<u>Percentage Correct</u>	<u>Grade 12</u>	<u>Grade 13</u>
90-99	1	4
80-89	4	5
70-79	2	5
60-69	4	4
50-59	7	-
40-49	-	-
30-39	-	-
20-29	1	1
n	19	19

TABLE A4.4

Discrimination Indices for the Items in the
 Reading Comprehension Part of the Test de
connaissance de la langue (anglais)
 (N = 470)

<u>Item</u>	<u>Discrimination Index^a</u>
1	47
2	45
3	62
4	59
5	68
6	62
7	49
8	56 ^b
9	-
10	53
11	54
12	49
13	46
14	63
15	50
16	58
17	61
18	56
19	59
20	57
Mean	55
S.D.	6

Note: Decimal points have been omitted.

^aThe biserial correlation between item scores and total test scores, uncorrected for the inclusion of the item in the test, is the index of item discrimination.

^bThis item was declared faulty in that it had two correct answers. It was excluded from subsequent analysis.

TABLE A4.5

Distribution Statistics and Information on the
Reliability and Standard Error of Measurement
of the Reading Comprehension Part of the
Test de connaissance de la langue (anglais)

Number of Examinees	470
Number of Test Items ^a	19
Mean ^b	11.95
S.D. ^b	4.17
Highest Score	19.00
Lowest Score ^c	-2.33
Reliability ^d	0.71
Standard Error of Measurement ^d	2.21

^aThe number of items in the test was 20 but one was declared faulty because it had two answers and was omitted.

^bThis statistic is for the distribution of corrected-for-guessing scores before any adjustment was made for the effect of absenteeism.

^cNegative scores arise because a correction for guessing was applied. Wrong answers were scored $-1/3$, and correct answers 1. Omitted questions were ignored in scoring.

^dEstimated from a formula due to Hoyt (1941).

TABLE A4.6

Data on the Reliability With Which the
Writing Exercise Was Scored

<u>Marker</u>	<u>Estimates of Reliability</u>	
	<u>Summary Score</u>	<u>Commentary Score</u>
1	.21	.71
2	.61	.88
3	.49	.85
4	.55	.61
5	.83	.83
6	.77	.90
Mean	.58	.80
S.D.	.20	.10

APPENDIX A5

TECHNICAL REPORT ON THE PHYSICS ACHIEVEMENT TEST

(This Appendix was prepared by John Hattie, a Graduate Student at the Ontario Institute for Studies in Education.)

1. DESCRIPTION OF TEST CONTENT

The test under consideration in this report is one form of the 1970 version of the Ontario Physics Achievement Test. This test was part of a battery called the Ontario Tests for Admission to College and University.

The Physics Achievement Test is based on the Grade Thirteen curriculum formulated by the Ontario Department of Education (Curriculum S.17C, 1967). The main theme in this curriculum is the wave-particle duality of radiation matter. The curriculum is divided into four units. Only the first three units are assessed by the test.

Unit I is entitled "Time, Space, and Motion", and is a general introduction to the concepts of time and space. It focusses on fundamentals of measurement, functions, motion along a straight line path, and motion in space. Unit II uses the study of light as a vehicle to show how physical models or theories are developed. It is called "Optics and Waves", and concerns the behaviour of light, the particle model of light, waves propagated in one and two dimensions, interference, and the light and wave model. Unit III is called "Mechanics" and deals with the dynamics of particles. Specifically, the topics covered are the

law of inertia and Newton's third law, motion in the earth's gravitational field, the universal law of gravitation, the solar system, momentum and the conservation of momentum, work and kinetic energy, and potential energy.

The Physics Achievement Test consists of 18 items on material in Unit I (items 1-9, 31-39), 19 items for Unit II (items 10-19, 40-48), and 23 items for Unit III (items 20-30, 49-60). The topics covered are: Graphical solutions (14 items: 1, 5, 6, 14, 19, 27, 28, 33, 34, 35, 49, 50, 56, 57); Ratio and Proportion (10 items: 2, 3, 17, 23, 31, 33, 38, 39, 42, 52); Abstractions (11 items: 1, 4, 22, 23, 30, 38, 39, 46, 52, 58, 59); Linear Kinematics (11 items: 1, 4, 5, 6, 10, 21, 22, 35, 37, 53, 54); Non-linear Kinematics (8 items: 7, 8, 9, 24, 25, 32, 38, 39); Dynamics (19 items: 20, 23, 26, 27, 28, 29, 30, 34, 36, 49, 50, 51, 52, 55, 56, 57, 58, 59, 60); Vectors (7 items: 5, 7, 8, 9, 24, 25, 26); Light (9 items: 11, 12, 14, 17, 40, 41, 46, 47, 48); and Waves (8 items: 13, 15, 16, 18, 19, 43, 44, 45).

The abilities tested by the items in this instrument are as follows:

- (a) Ability to demonstrate an understanding of basic scientific concepts and principles.

Students are required to demonstrate their understanding of facts and their ability to reason with them, rather than their ability merely to recognize facts.

- (b) Ability to apply scientific concepts and principles

To respond correctly to items measuring this ability, students must not only understand a principle, but they must also recognize how to apply it to many different situations, both in the physical world and in the laboratory. They are required to deduce specific inferences from broad generalizations and concepts.

(c) Ability to handle quantitative relationships.

This concerns reasoning in quantitative terms. That is, students must be able to understand the "how much" or the "how little" involved in scientific problems and be able to work out the numerical relationships between quantities.

(d) Ability to interpret cause-and-effect relationships

Students are asked to select the probable causes for happenings in nature from a list of principles and concepts related in other ways to the happenings. In addition, they are asked to predict the effect that altering one or more variables may have on a system.

(e) Ability to apply laboratory procedures and interpret experimental data.

It is expected that students who take this test will have had laboratory experience of a sort that familiarizes them with methods of formulating and attacking experimental problems. Since physics is an experimental science, students who have studied it should be familiar with the types of experimental results commonly obtained in physics and should know how to interpret such results.

Below are two examples of the items contained in the test:

1. Two sources, S_1 and S_2 , separated by distance d , illuminate a screen a distance L away. Which one of the following is a necessary condition for an observable stationary interference patterns to be produced on the screen?

(A) Radiations from S_1 and S_2 differ in wavelength.

(B) The frequency of source S_1 is very close to but not equal to the frequency of the source S_2 .

- (C) The distance d is large compared with the distance L .
- (D) One source is farther from the screen than the other source.
- (E) There is a constant phase difference between source S_1 and source S_2 .

Answer: To answer this item it is necessary to realize that, assuming the frequencies of the sources S_1 and S_2 to be identical, there must be a fixed phase relation between the sources in order to produce a stationary interference pattern. Hence the answer is (E).

2. A steel ball of mass 0.1 kilogram is fired due west at a speed of 90 meters per second. After collision with another steel ball it is moving at 120 meters per second due south. The magnitude of the change of its momentum is:

- (A) 21 kg-m/sec
(B) 3 kg-m/sec²
(C) 3 kg-m/sec
(D) 15 kg-m/sec
(E) 15 kg-m/sec²

Answer: To answer this question correctly it is necessary to find the difference between the initial or final vectors with the aid of Pythagorean relation. The correct units for momentum must also be considered. The answer is statement (D). The test was translated from English into French for administration to Francophone students.

2. TEST APPRAISAL

Test Appraisal Inventories were sent to the teachers of Grade Thirteen physics in the 67 secondary schools, both Anglophone and Francophone, in the study. Teachers in 62 schools responded. Sixty-two English language, and 11 French language teachers responded. University professors teaching first year courses in 11 Ontario universities were also sent the inventory. Fifteen English language and 6 French language professors completed it. The secondary school teachers constitute a reasonably representative sample of physics teachers in the province. The university professors however, are not a probability sample of university professors since names were selected quite arbitrarily and a number of those who were asked to respond did not reply. It must also be remembered that as the number of postsecondary teachers is small, the precision of results for this group is not high.

2.1 Coded Responses to the Appraisal Inventory

The Test Appraisal Inventory aimed at judging the kind of knowledge required to answer the questions, and at assessing whether the content of the items concerned knowledge that students know or should know. The secondary school teachers were asked to consider each item in terms of:

A. Old knowledge that students should have on entry to the course:

A1. This knowledge is not reviewed in the course.

A2. This knowledge is reviewed in the course.

B. New knowledge that all students are expected to learn in the course.

C. New knowledge that some students are expected to learn in the course:

C1. Only 1 per cent to 25 per cent of students should learn this.

C2. Only 26 per cent to 50 per cent of students should learn this.

C3. Only 51 per cent to 75 per cent of students should learn this.

C4. More than 75 per cent but not all students should learn this.

D. New knowledge that no student is expected to learn.

The university professors assessed each item in terms of the following categories:

A. Old knowledge that students should have on entry to the course:

A1. This knowledge is not reviewed in the course.

A2. This knowledge is reviewed in the course.

B. New knowledge that all students are expected to learn in the course.

C. Other.

The instructions for the Test Appraisal Inventory received some comment. One teacher had difficulty distinguishing between knowledge that is reviewed and new knowledge that all students are expected to learn in the course. One teacher asked whether or not new knowledge that some students will be taught was the same

as new knowledge some students will learn. Another queried whether emphasis was on material taught in enough depth so that the student could attempt the question, or on an estimate of what the student would actually do with the question in the test situation. One teacher asked whether the questions were to be classified according to what he hoped to teach if he had time, or according to what he actually had taught. Finally, one teacher noted that if he taught the content, he would expect 95 per cent to learn that material.

The frequency tabulations for each category across each item are presented in Table A5.1. In an attempt to discover which items were thought to test knowledge that students at the interface should have, the frequency of secondary school appraisals falling in categories A1, A2, B, and C4 were accumulated for each item. (These reflect judgments that 75 per cent or more of students at the end of Grade Thirteen should have acquired the knowledge to answer the item correctly.) Also, the frequency of university appraisals falling in categories A1 and A2 were accumulated. (These reflect judgments that students who enter a first year university course in physics should already have acquired the knowledge tested by the item.) These accumulations are reported in Table A5.1 as Total 1. Total 2 constitutes the sum of the remaining appraisal categories.

The mean Total 1 percentage for English language secondary school teachers was 90.30 (standard deviation = 5.91), for French language secondary school teachers 83.46 (s.d. = 1.89), for English language university professors 56.89 (s.d. = 3.27), and for French language university professors 60.83 (s.d. = 1.41).

For the English language secondary school teachers, all but three items were judged to demand knowledge that more than 75 per cent of the appraisers said that 75 per cent or more of students at the interface should have. The three items not included were item 27 which concerns dynamics and a graphical solution; item 38 which concerns ratios and proportions,

abstractions, and non-linear kinematics; and item 39 which concerns non-linear kinematics and ratios and proportions. For the French language teachers there were 11 items that were classified as requiring knowledge that less than 75 per cent of appraisers said that 75 per cent or more of the students should have. Of these 11 items, 5 concerned Unit I material, 3 concerned Unit II material, and 3 concerned Unit III material.

For the university professors, a very low proportion of items met the criterion of testing knowledge that 75 per cent or more of the appraisers indicated students should have on entry to university. There were only 14 items which more than 75 per cent of the English language professors said tested knowledge the students should have on entry to a university course. Of the 46 items not included, 7 concerned Unit I material, 18 concerned Unit II material, and the other 21 concerned Unit III material. For the French language professors, there were only 17 items which more than 75 per cent claimed tested old knowledge. Of the 43 items not included, 7 concerned Unit I material, 18 concerned Unit II material, and 18 concerned Unit III material.

These results from the appraisal of the Physics Achievement Test/Test de rendement en physique suggest the conclusion that most secondary school teachers regarded the test as appropriate for students at the interface. University instructors, on the other hand, tended to view the test more negatively, most finding relatively few items that would test knowledge expected of entering first year students.

2.2 Written Responses to the Appraisal Inventory

There was an opportunity on the appraisal form for written comments about the test. Among the secondary school teachers there were three appraisers who responded that the physics test was suitable for Grade Thirteen students and that the questions were well written. There was, however, at least one expression of concern over the fact that material in the test could not be

covered for each of the following reasons: shortage of time--a result of the Toronto teachers' strike; semestering, which put spring semester students behind students in schools not on a semester schedule; and insufficient time, due to the fact that fewer hours are assigned to the Physical Sciences Study Committee course here in Ontario than in the United States, where it originated. Further, one teacher commented that when the goals of the course were set a few years ago, only the top quarter of the students studied physics, whereas more than one-half of Grade Thirteen students now take the same course. Because of this, the ability range of students is wider than it formerly was; hence the teaching techniques required to cope with students nowadays are quite different from the techniques required in the past. These reasons could account for a number of teachers claiming that some topics covered in the test were used as enrichment material only.

Seven secondary school teachers expressed concern about the conditions under which the test was administered. They claimed that the students were not adequately forewarned about the test and there was little incentive for students to perform as well as they could.

For many of the above reasons it was noted by two teachers that test results should not be compared with other tests administered on earlier occasions.

One teacher commented that an all or nothing marking system disadvantaged the students, particularly as a review was not undertaken. His argument was that many students might remember the process but not the exact formula. Although these individuals should be able to eliminate some incorrect responses, if they chose the wrong answer they would receive no credit. (It should be pointed out in response to this comment that students writing the physics test were told that guessing would be penalized; the application of a correction for guessing has the effect of crediting partial knowledge.)

A pot-pourri of additional general comments were made by secondary school teachers. Notation was mentioned--units in the test were not Systeme Internationale, and teachers used d, not s, for distance. It was claimed that the ability to memorize formulae and the topic of waves received too much emphasis in the test. Some teachers felt that some topics which should have been tested were not; topics mentioned included qualitative electromagnetism; modern models of light charge and matter; units of measure and scientific notation; electrostatics; nuclear and atomic physics; vectors or components; waves; particle theory of light and the nature of light. Other appraisers noted that there were few items on electricity and atomic structure. (It is obvious that these teachers failed to remember that Unit IV of the Grade Thirteen curriculum was not intended to be assessed. Unit IV is entitled "Electricity and Atomic Structure", and covers electromagnetism, electrostatics, and nuclear and atomic physics.)

The secondary school teachers made other comments about specific items. The most frequent comment was that a particular item was "not covered" in the course. This comment was made 111 times in all and there were only 15 questions that did not receive this comment at least once. Seventeen items were considered to be enrichment material by one or two teachers. Five questions were regarded by up to three teachers as testing mathematics, not physics (items 1, 2, 3, 31, and 33). Two teachers claimed students could have answered items 12 and 13 correctly if there had been time to review formulae. Four teachers each found an item that had, in their opinion, incorrect answers or insufficient information (items 14, 27, 35, and 40). (This contention is, of course, disputed by the experts who constructed the test and, presumably, also by the vast majority of appraisers who did not make a similar comment.) There was a complaint about the fact that three items could only be answered correctly if previous questions had been answered correctly (item 16 depended on 15; 35 and 36 depended on 34; and 39 depended on 38).

Some items elicited comments reflecting differences in teaching styles. For example, one teacher saw the wording of item 19 concerning waves as being ambiguous, while another was concerned that the details necessary for answering the question were soon forgotten; another claimed that numbers should have been used, and yet another claimed that the use of graphs was misleading. The use of graphs in five questions (items 19, 27, 28, 30, and 35) received negative comments from five teachers, and 20 teachers each found one item that required a degree of abstraction deemed to be beyond the Grade Thirteen student. One teacher claimed that the picture accompanying item 46 on wavelengths was ambiguous and another teacher contended that item 48 on waves required awareness and recall of an actual experiment.

These comments by the secondary school teachers centred primarily around the lack of time for review and preparation for this test and lack of time to teach the curriculum. The other comments concerned specific details of the test.

The university professors also made commendatory and critical comments. The type of physics course each professor taught seemed to influence his comment. One said, for example, that the test was good for students taking physics for engineering, while another, who taught physics for music, claimed the test was inappropriate. Another noted that much of the material that concerned first principles was automatically reviewed when a higher level was being taught. It was reported by one professor that questions given to university students tended to be slightly more quantitative than those on the test.

The number of comments from the university professors on specific items was small. Some concerned items that were seen by the professors as pertaining to material not covered in a first year course. Altogether, there were 20 occasions when the comment "not covered in the course" was placed by an item. Specifically mentioned as not covered were items relating to optics, variable forces, relative acceleration, nodal lines and

wave motion. Two items, (number 42 and number 50) were identified--each by a different single appraiser--as not having a correct answer.

Overall, the appraisers' comments reflected a concern about material not covered in their courses.

3. TECHNICAL ISSUES

3.1 Scoring

Each item was scored 1 if correct, -0.125 if incorrect, and 0 if omitted. Prior to taking the test, students were told that when the test was scored, a percentage of the wrong answers would be subtracted from the number of right answers as a correction for haphazard guessing. However, if they had sufficient knowledge of the problem to eliminate one or more choices as definitely wrong, they were informed that it would be to their advantage to guess among the remaining choices.

3.2 Difficulty

The mean uncorrected test score for the 641 English students who wrote the test was 19.56, with a standard deviation of 8.90. The range of scores was from 2 to 53. For the 82 Francophone students, the mean uncorrected test score was 14.13, with a standard deviation of 6.81, and a range from 0 to 41.

An index of difficulty was estimated for each item in the test as follows. The percentage of students answering an item correctly in a given school was obtained. This percentage was then averaged over all the Anglophone (Francophone) schools to obtain the Anglophone (Francophone) index of difficulty for an item.

The mean difficulty indices for the items in various subgroups are reported in Table A5.2. The overall mean difficulty index was 33 (standard deviation = 16) for the Anglophone students. For the Francophone students the mean was 24 (standard deviation = 16). Both the low mean score on this test and the low mean index of item difficulty point to the fact that the test was very difficult for the students who wrote it. In tests designed to spread students over the possible range of scores, it is common to strive for an average item difficulty at the point midway between perfect performance for everyone on an item and performance at a level that could be expected from pure guessing. For a test composed of five option multiple-choice items, this ideal level of difficulty is approximately 60 per cent of correct answers. The Physics Achievement Test is obviously very much more difficult than this.

3.3 Appraisers Expectations and Difficulty

To assess whether the appraisers' judgements about the items were related to item difficulty, a coefficient of correlation was computed between the two variables. These correlations for each set of appraisers are presented in Table A5.3. The estimates of difficulty that were used to compute the coefficients for Anglophone appraisers were from Anglophone students; the difficulty estimates used to compute the coefficients for Francophone appraisers were from Francophone students. The relationship between the appraisers' comments and item difficulty is similar, and moderately high; correlations of this size reflect the fact that 10 per cent of the variance in item difficulty can be predicted from the appraisers' judgments.

3.4 Discrimination

A measure of the extent to which an item discriminates is the biserial correlation coefficient between scores on an item and scores on the total test. In tests designed to spread students

over the range of scores, it is considered desirable to have biserial correlation coefficients of 0.3 or higher. The biserial correlation coefficients were computed for the items in this test. The indices were computed separately for Anglophone and Francophone students. A frequency distribution of these coefficients is reported in Table A5.4. Seven items in the test for English students, and 18 items in the test for French students had substandard indices of discrimination. On the whole, however, the items discriminated at an acceptably high level; the mean discrimination index for English students was 0.45 (s.d. = 0.13), and for French students the mean was 0.42 (s.d. = 0.23).

3.5 Reliability

The Hoyt estimate of reliability for uncorrected scores for Anglophone students was 0.87, with a standard error of measurement of 3.21, and for Francophone students was 0.81 a standard error of measurement of 2.93. (See Hoyt, 1941 for the formulae used to compute this coefficient of reliability and standard error of measurement.) These results compare favourably with the KR(20) estimate of the reliability of uncorrected scores (0.86) and the associated standard error of measurement (3.3) reported in the 1970 test manual.

3.6 Speededness

Using criteria formulated by Educational Testing Service, a test is considered unspeeded if at least 80 per cent of those writing reach the last item, and all candidates reach the three-quarter mark. For this administration of the Physics Achievement Test/
Test de rendement en physique, statistics on the percentage of students not reaching a given item are presented in Table A5.5. For the English students, 3 per cent of the students did not reach the three-quarter mark and 37 per cent did not reach the last item. For the French students, 4 per cent did not reach the

three-quarter mark, and 59 per cent did not reach the last item. Hence the test was somewhat speeded. It should be noted that this test was very difficult for most candidates and it may be that the items near the end of the test were reached but not attempted.

REFERENCES

Hoyt, C. Test reliability estimated by analysis of variance. Psychometrika, 1941, 6, pp. 153-160.

TABLE A5.1

Frequency Tabulations of the Responses of Appraisers to Items
on the Physics Achievement Test/
Test de rendement en physique

Item	Teachers of	A1	A2	B	C4	C3	C2	C1	D	Total 1	Total 2
1	Sec	8	41	11	0	0	0	0	2	60	2
	F	1	8	3	0	0	1	0	0	12	1
	Univ	11	4	0	0	0	0	0	0	15	0
	F	3	3	0	0	0	0	0	0	6	0
2	Sec	9	27	20	0	0	2	0	4	56	6
	F	6	6	0	0	1	0	0	0	12	1
	Univ	11	4	0	0	0	0	0	0	15	0
	F	5	1	0	0	0	0	0	0	6	0
3	Sec	4	16	30	0	0	2	2	8	50	12
	F	1	3	5	0	0	0	1	3	9	4
	Univ	8	3	1	0	0	0	0	2	11	3
	F	5	0	1	0	0	0	0	0	5	1
4	Sec	1	29	29	1	0	1	0	0	60	1
	F	0	7	6	0	0	0	0	0	13	1
	Univ	1	11	2	0	0	0	0	1	12	0
	F	1	5	0	0	0	0	0	0	6	3
5	Sec	1	38	22	1	0	0	0	0	62	0
	F	0	6	7	0	0	0	0	0	6	0
	Univ	2	11	1	0	0	0	0	1	13	7
	F	1	5	0	0	0	0	0	0	6	2
6	Sec	2	36	23	0	1	0	0	0	61	0
	F	0	6	7	0	0	0	0	0	13	1
	Univ	3	10	1	0	0	0	0	1	13	0
	F	1	5	0	0	0	0	0	0	6	0

TABLE A5.1 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>D</u>	<u>Total 1</u>	<u>Total 2</u>
7	Sec	0	14	46	2	0	0	0	0	62	0
	Univ	0	2	10	1	0	0	0	0	13	0
8	Sec	4	8	2	0	0	0	0	1	12	3
	Univ	2	3	1	0	0	0	0	0	5	1
9	Sec	0	6	53	2	1	0	0	0	61	1
	Univ	0	3	9	0	1	0	0	0	12	1
10	Sec	3	8	3	0	0	0	0	1	11	4
	Univ	2	3	1	0	0	0	0	0	5	1
11	Sec	0	4	54	3	1	0	0	0	61	1
	Univ	1	0	9	0	0	2	1	0	10	3
12	Sec	2	9	3	0	0	0	0	1	11	4
	Univ	1	4	1	0	0	0	0	0	5	1
13	Sec	8	20	25	3	2	2	1	1	56	6
	Univ	2	4	7	0	0	0	0	0	13	0
14	Sec	5	6	1	0	0	0	0	3	11	4
	Univ	2	2	1	0	0	0	0	1	4	2
15	Sec	3	28	26	0	1	0	1	3	57	5
	Univ	1	6	4	1	0	0	0	0	12	1
16	Sec	3	5	3	0	0	0	0	4	8	7
	Univ	1	3	1	0	0	0	0	1	4	2
17	Sec	3	27	26	0	1	0	2	3	56	6
	Univ	1	5	6	0	0	0	1	0	12	1
18	Sec	3	5	3	0	0	0	0	4	8	7
	Univ	1	2	2	0	0	0	0	1	4	2
19	Sec	8	28	19	0	1	0	1	5	55	7
	Univ	4	4	4	0	0	0	1	0	12	1
20	Sec	1	6	5	0	0	0	0	3	7	8
	Univ	0	3	1	0	0	0	0	0	3	3

TABLE A5.1 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>D</u>	<u>Total 1</u>	<u>Total 2</u>
14	Sec	6	19	30	1	1	0	0	5	56	6
		2	4	5	0	0	0	2	0	11	2
	Univ	1	6	2	0	0	0	0	6	7	8
		0	3	2	0	0	0	0	1	3	3
15	Sec	19	26	15	0	1	0	0	1	60	2
		4	4	3	0	0	0	1	1	11	2
	Univ	2	6	4	0	0	0	0	3	8	7
		1	1	2	0	0	0	0	2	2	4
16	Sec	16	30	15	0	0	0	0	1	61	1
		4	5	2	0	0	0	1	1	11	2
	Univ	3	5	4	0	0	0	0	3	8	7
		1	1	2	0	0	0	0	2	4	2
17	Sec	2	1	5	2	0	1	0	1	60	2
		1	0	9	1	0	1	1	0	11	2
	Univ	1	4	4	0	0	0	0	6	5	10
		0	3	0	0	0	0	0	3	3	3
18	Sec	2	2	5	3	1	1	0	1	59	1
		1	1	9	1	0	0	1	0	12	10
	Univ	1	4	6	0	0	0	0	4	5	3
		1	2	0	0	0	0	0	3	3	10
19	Sec	2	2	4	4	4	2	1	0	52	10
		2	0	7	1	0	0	2	6	10	2
	Univ	1	2	6	0	0	0	0	3	3	12
		0	3	0	0	0	0	0	3	3	3
20	Sec	0	6	5	2	1	0	0	0	61	1
		0	2	9	1	0	0	1	0	12	1
	Univ	1	9	3	0	0	0	0	2	10	5
		1	3	2	0	0	0	0	0	4	2

TABLE A5.1 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>D</u>	<u>Total 1</u>	<u>Total 2</u>
21	Sec	1	20	41	0	0	0	0	0	62	0
	F	0	5	8	0	0	0	0	0	13	0
	Univ	2	10	2	0	0	0	0	1	12	3
	F	1	4	1	0	0	0	0	0	5	1
22	Sec	2	16	42	0	0	1	0	1	60	2
	F	0	6	7	0	0	0	0	0	13	0
	Univ	4	9	1	0	0	0	0	1	13	2
	F	1	4	1	0	0	0	0	0	5	1
23	Sec	0	3	52	5	1	0	1	0	60	2
	F	0	0	11	1	0	0	1	0	12	1
	Univ	0	10	4	0	0	0	0	1	10	5
	F	0	3	3	0	0	0	0	0	3	3
24	Sec	1	3	43	4	1	0	3	6	51	10
	F	0	0	9	1	0	0	2	1	10	3
	Univ	0	7	5	0	0	0	0	3	7	8
	F	1	2	0	0	0	0	0	3	3	3
25	Sec	1	10	39	3	1	0	3	5	53	9
	F	1	2	8	0	0	0	2	1	10	3
	Univ	0	7	4	0	0	0	0	4	7	8
	F	1	2	1	0	0	0	0	2	3	3
26	Sec	0	0	54	4	2	1	0	1	58	4
	F	0	0	12	0	0	0	1	0	12	1
	Univ	1	5	7	0	0	0	0	2	6	9
	F	0	3	3	0	0	0	0	0	3	3
27	Sec	0	2	35	2	6	7	5	5	39	23
	F	0	4	6	1	0	0	0	2	11	2
	Univ	0	4	7	0	0	0	0	4	4	11
	F	0	0	4	0	0	0	0	2	0	6

TABLE A5.1 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>D</u>	<u>Total 1</u>	<u>Total 2</u>
28	Sec	0	13	43	3	2	1	0	0	59	3
	F	1	4	7	1	0	0	0	0	13	0
	Univ	1	7	6	0	0	0	0	1	8	7
	F	0	2	4	0	0	0	0	0	2	4
29	Sec	1	27	31	1	1	0	0	0	60	1
	F	2	4	5	0	0	1	0	1	11	2
	Univ	2	6	6	0	0	0	0	1	8	7
	F	0	4	2	0	0	0	0	0	4	2
30	Sec	0	9	41	2	7	1	2	0	52	10
	F	2	0	7	0	0	2	1	1	9	4
	Univ	0	9	5	0	0	0	0	1	9	6
	F	0	2	4	0	0	0	0	0	2	4
31	Sec	12	24	16	1	1	0	1	7	53	9
	F	4	4	3	1	0	0	1	0	12	1
	Univ	12	3	0	0	0	0	0	0	15	0
	F	6	0	0	0	0	0	0	0	6	0
32	Sec	1	6	42	3	1	1	1	8	52	10
	F	1	0	7	0	1	0	0	3	8	5
	Univ	10	2	1	0	0	0	0	2	12	3
	F	4	0	1	0	0	0	0	1	4	2
33	Sec	9	13	24	2	3	5	1	5	48	14
	F	4	1	4	0	1	0	1	1	9	3
	Univ	10	2	2	0	0	0	0	1	12	3
	F	4	1	1	0	0	0	0	0	5	1
34	Sec	1	23	37	1	0	0	0	0	62	0
	F	0	6	7	0	0	0	0	0	13	0
	Univ	4	8	2	0	0	0	0	1	12	3
	F	1	4	1	0	0	0	0	0	5	1

TABLE A5.1 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>D</u>	<u>Total 1</u>	<u>Total 2</u>
35	Sec	0	23	37	1	0	0	1	0	61	1
	F	0	5	8	0	0	0	0	0	13	0
	Univ	3	9	2	0	0	0	0	1	12	3
	F	1	4	1	0	0	0	0	0	5	1
36	Sec	0	4	55	1	2	0	0	0	60	2
	F	0	1	9	0	1	0	2	0	10	3
	Univ	1	4	8	0	0	0	0	2	5	10
	F	1	2	3	0	0	0	0	0	3	3
37	Sec	0	7	49	3	0	1	1	1	59	3
	F	0	1	7	1	0	1	1	2	9	4
	Univ	1	6	7	0	0	0	0	1	7	8
	F	0	4	2	0	0	0	0	0	4	2
38	Sec	10	5	23	0	2	3	3	16	38	24
	F	1	1	7	0	0	1	2	1	9	4
	Univ	7	3	3	0	0	0	0	2	10	5
	F	4	1	1	0	0	0	0	0	5	1
39	Sec	0	1	34	2	1	4	4	16	37	25
	F	0	0	7	1	1	1	2	1	8	5
	Univ	3	5	4	0	0	0	0	3	8	7
	F	1	2	3	0	0	0	0	0	3	3
40	Sec	4	9	44	1	1	0	0	3	58	4
	F	3	3	6	0	0	0	1	0	12	1
	Univ	3	4	3	0	0	0	0	5	7	8
	F	2	1	2	0	0	0	0	1	3	3
41	Sec	11	18	24	1	2	1	2	3	54	8
	F	5	2	5	0	0	0	1	0	12	1
	Univ	1	4	5	0	0	0	0	5	5	5
	F	3	0	1	0	0	0	0	2	3	3

TABLE A5.1 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>D</u>	<u>Total 1</u>	<u>Total 2</u>
42	Sec	1	4	43	3	2	4	2	3	51	11
		1	1	8	0	0	0	1	2	10	3
	Univ	4	1	3	0	0	0	0	7	5	10
		2	1	2	0	0	0	0	1	3	3
43	Sec	2	31	27	0	0	0	0	2	60	2
		3	4	5	0	0	0	1	0	12	1
	Univ	5	2	5	0	0	0	0	3	7	8
		1	1	2	0	0	0	0	2	2	4
44	Sec	2	7	35	2	3	1	5	6	46	15
		3	3	4	0	0	0	1	2	10	3
	Univ	1	6	5	0	0	0	0	3	7	8
		1	1	2	0	0	0	0	2	2	4
45	Sec	3	5	48	2	1	0	1	2	58	4
		4	1	5	0	1	0	2	0	10	3
	Univ	7	0	2	0	0	0	0	6	7	8
		3	1	0	0	0	0	0	2	4	2
46	Sec	2	2	41	4	3	1	4	5	49	13
		0	0	5	1	0	1	2	4	6	7
	Univ	1	1	7	0	0	0	0	6	2	13
		0	2	1	0	0	0	0	3	2	4
47	Sec	2	1	42	5	3	1	3	5	50	12
		1	0	8	0	0	2	1	1	9	4
	Univ	0	4	5	0	0	0	0	6	4	11
		0	3	0	0	0	0	0	3	3	3
48	Sec	2	1	41	5	3	2	3	5	49	13
		0	0	7	0	1	1	1	3	7	6
	Univ	0	2	7	0	0	0	0	6	2	13
		0	2	1	0	0	0	0	3	2	4

TABLE A5.1 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>D</u>	<u>Total 1</u>	<u>Total 2</u>
49	Sec	1	23	36	2	0	0	0	0	62	0
		0	4	8	0	0	1	0	0	12	1
	Univ	2	8	3	0	0	0	0	2	10	5
		1	4	1	0	0	0	0	0	5	1
50	Sec	0	3	52	2	2	2	1	0	57	5
		0	1	9	0	0	1	0	2	10	3
	Univ	2	4	7	0	0	0	0	2	6	9
		1	2	3	0	0	0	0	0	3	3
51	Sec	3	17	40	0	1	0	0	1	60	2
		1	6	6	0	0	0	0	0	13	0
	Univ	5	5	4	0	0	0	0	1	10	5
		2	3	1	0	0	0	0	0	5	1
52	Sec	0	23	35	1	1	0	0	2	59	3
		0	5	8	0	0	0	0	0	13	0
	Univ	4	7	3	0	0	0	0	1	11	4
		3	2	1	0	0	0	0	0	5	1
53	Sec	0	15	44	2	1	0	0	0	61	1
		2	2	9	0	0	0	0	0	13	1
	Univ	2	10	2	0	0	0	0	1	12	0
		1	3	2	0	0	0	0	0	4	3
54	Sec	0	5	53	3	1	0	0	0	61	2
		0	0	13	0	0	0	0	0	13	1
	Univ	3	8	3	0	0	0	0	1	11	0
		1	3	2	0	0	0	0	0	4	4
55	Sec	0	2	53	4	0	1	1	1	59	2
		0	0	10	0	0	2	1	0	10	3
	Univ	1	5	7	0	0	0	0	2	6	9
		1	2	3	0	0	0	0	0	3	3

TABLE A5.1 (continued)

Item	Teachers of	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>D</u>	<u>Total 1</u>	<u>Total 2</u>
56	Sec	0	4	52	1	2	1	0	2	57	5
	F	0	1	9	0	1	0	2	0	10	3
	Univ	0	7	5	0	0	0	0	3	7	8
	F	0	2	3	0	0	0	0	1	2	4
57	Sec	0	5	47	4	1	2	2	1	56	6
	F	0	1	10	0	1	0	1	0	11	2
	Univ	0	6	7	0	0	0	0	2	6	9
	F	0	2	3	0	0	0	0	1	2	4
58	Sec	0	1	50	3	3	3	1	1	54	8
	F	0	1	7	0	0	1	1	3	8	5
	Univ	0	7	6	0	0	0	0	2	7	8
	F	0	1	5	0	0	0	0	0	1	5
59	Sec	0	2	47	2	2	2	1	6	51	11
	F	0	1	6	0	1	1	1	3	7	6
	Univ	2	2	8	0	0	0	0	3	4	11
	F	0	1	4	0	0	0	0	1	1	5
60	Sec	0	3	51	3	1	1	3	0	57	5
	F	0	1	10	0	1	0	0	1	11	2
	Univ	0	6	8	0	0	0	0	1	6	9
	F	0	3	3	0	0	0	0	0	3	3

Note: See section 2 (TEST APPRAISAL) of this Technical Report for explanation of A1 - D and of Total 1 and Total 2. The number of appraisals by each category of teachers varies from item to item because of omitted responses on the part of the appraisers.

TABLE A5.2

Mean Percentage of Students Responding Correctly to the
Items in Each of Several Categories of the
Physics Achievement Test / Test de rendement en physique

<u>Type</u>	<u>No. of Items</u>	<u>English</u>	<u>French</u>
Unit I	18	38	30
Unit II	19	29	23
Unit III	23	27	18
Graphs	14	32	22
Ratio and Proportion	10	31	24
Abstractions	11	25	19
Linear Kinematics	11	31	22
Non-Linear Kinematics	8	38	34
Dynamics	19	30	20
Vectors	7	49	42
Light	9	25	21
Waves	8	35	27

TABLE A5.3
Correlations Between Item Difficulty and
Total 1 of Table A5.1

		Item Difficulty	
		<u>English</u>	<u>French</u>
Secondary	E	0.39	0.16
	F	0.40	0.31
University	E	0.34	0.28
	F	0.30	0.26

TABLE A5.4

Distribution of Biserial Correlations
for Items in the Physics Achievement Test/
Test de rendement en physique

<u>Correlation</u>	<u>English</u>	<u>French</u>
-0.20 - 0.00	0	2
0.00 - 0.09	1	2
0.10 - 0.19	0	8
0.20 - 0.29	6	6
0.30 - 0.39	9	9
0.40 - 0.49	18	9
0.50 - 0.59	17	12
0.60 - 0.69	9	6
0.70 - 0.79	0	3
0.80 - 0.89	0	2
0.90 - 0.99	0	1

TABLE A5.5

Mean Difficulty, Appraisers' Total 1 Percentages, and Number of Students for Whom the Given Item was the First "Not Reached" for English (N=641) and French (N=82) Students

Item	Difficulty		Secondary		University		English		French		English		French	
	Eng	Fr	English	French	English	French	NR	Cum	NR	Cum	NR	Cum	NR	Cum
1	72	68	97	100	100	100	0	0	1	1	0	0	1	1
2	54	40	90	92	100	100	0	0	0	0	0	0	0	1
3	17	16	81	69	73	83	0	0	0	0	0	0	0	1
4	28	21	98	100	80	100	0	0	0	0	0	0	0	1
5	39	23	100	86	87	100	0	0	0	0	0	0	0	1
6	31	15	98	100	87	100	0	0	0	0	0	0	0	1
7	71	66	100	100	80	83	0	0	0	0	0	0	0	1
8	69	60	98	92	73	83	0	0	0	0	0	0	0	1
9	45	42	98	77	73	83	0	0	0	0	0	0	0	1
10	35	18	90	100	73	67	0	0	0	0	0	0	0	1
11	37	20	92	92	53	67	0	0	0	0	0	0	0	1
12	22	18	90	92	53	83	0	0	0	0	0	0	0	1
13	37	33	89	92	47	50	0	0	0	0	0	0	0	1
14	21	21	90	85	53	67	0	0	0	0	0	0	0	1
15	65	55	97	85	53	33	0	0	0	0	0	0	0	1
16	23	15	98	85	53	67	0	0	0	0	0	0	0	1
17	57	52	97	95	33	50	0	0	0	0	0	0	0	1
18	21	7	97	92	33	50	0	0	0	0	0	0	0	1
19	27	17	84	83	20	50	0	0	0	0	0	0	0	1
20	54	43	98	92	67	27	0	0	0	0	0	0	0	1

TABLE A5.5 (continued)

Item	Difficulty		Secondary		University		English		English		French	
	Eng	Fr	English	French	English	French	English	French	NR	Cum	NR	Cum
21	23	13	100	100	80	83	0	0	0	1	0	1
22	13	5	97	100	87	83	0	0	0	1	0	1
23	33	17	97	92	67	50	0	0	0	1	0	1
24	33	20	97	92	67	50	0	0	0	1	0	1
25	32	33	85	77	47	50	0	0	0	1	0	1
26	59	62	94	92	40	50	0	0	0	1	0	1
27	23	18	63	85	27	0	0	0	0	1	0	1
28	56	27	95	100	53	33	0	0	0	1	0	1
29	37	20	98	85	53	67	0	0	0	1	0	1
30	21	11	84	69	60	33	0	0	0	1	0	1
31	36	20	85	92	100	100	0	0	0	1	0	1
32	18	18	84	62	80	67	1	1	1	1	0	1
33	32	21	77	75	80	33	0	1	0	1	0	1
34	52	33	100	100	80	33	0	1	0	1	0	1
35	51	37	98	100	80	33	0	1	2	3	0	3
36	31	18	97	77	33	50	0	1	0	3	0	3
37	21	13	95	69	47	67	0	1	0	3	0	3
38	31	28	61	69	66	83	0	1	0	3	0	3
39	17	21	60	62	53	50	0	1	0	3	0	3
40	20	17	94	92	47	20	0	1	0	3	0	3
41	22	18	87	92	50	50	1	2	0	3	0	3
42	15	5	82	77	33	50	0	2	0	3	0	3
43	68	56	97	92	47	33	0	2	0	3	0	3
44	25	18	74	77	47	33	0	2	1	4	1	4
45	25	12	94	77	47	67	1	3	0	4	0	4
46	12	9	79	46	13	33	1	4	0	4	0	4
47	36	24	81	69	27	50	0	4	1	5	0	5
48	17	10	79	54	13	33	0	4	0	5	0	5
49	26	17	100	92	67	83	0	4	0	6	1	6

TABLE A5.5 (continued)

Item	Difficulty		Secondary		University		English		French	
	Eng	Fr	English	French	English	French	English	French	English	French
50	20	13	92	100	40	50	2	6	0	6
51	37	34	97	100	67	83	1	7	3	9
52	36	21	95	100	73	83	0	7	3	12
53	26	13	98	100	80	67	1	8	0	12
54	27	10	98	100	73	27	3	14	5	17
55	32	20	95	77	40	50	2	13	1	18
56	9	5	92	77	47	33	3	16	5	23
57	15	6	90	85	40	33	1	17	1	24
58	9	1	87	62	47	17	2	19	6	30
59	18	6	82	54	27	17	5	24	4	34
60	32	15	92	85	40	50	13	37	25	59
Mean	33	24	90	83	57	61				
S.D.	16	16	6	2	3	1				

^aNR = Not Reached; Cum = Cumulative percentage of students who did not reach the item.

APPENDIX A6

TECHNICAL REPORT

MATHEMATICS ACHIEVEMENT TEST

The Mathematics Achievement Test was administered to 1,404 students potentially eligible in June 1976 for a Secondary School Honour Graduation Diploma (SSHGD). These students were drawn from 52 of the 53 Anglophone schools involved in the study (the remaining school contains no SSHGD-level students). The population tested included students enrolled during the school year 1975-1976 in both Calculus and Relations and Functions.

The test had originally been administered as Part I of the Ontario Mathematics Achievement Test, Form OB, in May 1968. It consisted of 36 multiple-choice 5-option items; the time allowed was 60 minutes. The description of the test content which follows is excerpted from the Technical Report on the original administration, written by Dorothy M. Horn and Wilfred G. Fletcher:

The items were originally written by Grade 13 mathematics teachers in the Province of Ontario and by members of the Ontario Mathematics Achievement Test Committee. This committee reviewed all items, deleting some and revising others. They selected items for tryout tests, given in the Fall, 1967, and later made the selection of items for the test, Form OB.

The items tested were distributed among the following three categories:

I. Grade 13, Analysis

- A. Function as a mapping
- B. Second degree relations in the plane
- C. Trigonometry
- D. Transformations in the plane
- E. Slopes and simple derivatives
- F. Applications of differentiation

II. Basic principles from earlier grades related to topics such as:

- A. Linear and quadratic functions
- B. Exponents and logarithms
- C. Circles
- D. Sequences and series

III. Miscellaneous (items which cannot be identified clearly with a particular grade level or a particular topic in the Courses of Study for Grades 10, 11, 12 or 13).

The taxonomy classifications were:

- A. Knowledge and information: definitions, notations, concepts;
- B. Techniques and skills: solutions;

C. Translation of data into symbols or schema and vice versa;

D. Comprehension: capacity to analyze problems and to follow reasoning;

E. Inventiveness: reasoning creatively in mathematics.

Some further comments should be made concerning the test content. The test was, of course, developed at a time when Senior Division Mathematics courses had a structure quite different from the present one. Consequently, the fit between test content and curriculum content on this administration was considerably less good than had been the case in 1968. The test provides a comprehensive coverage of Relations and Functions, and of parts of Calculus, but has no items dealing with such areas as solution of differential equations, polar co-ordinates, and complex numbers. There is no coverage of Algebra. Since it proved impracticable to split the test into two subtests dealing respectively with Relations and Functions and Calculus, so that students in these courses could be tested separately, the decision was taken to administer the test only to students enrolled in both these courses.

1. DIFFICULTY

Table A6.1 presents difficulty indices for all test items for both the original test administration (1968) and the present administration.

Considerable caution should be observed in drawing any conclusions regarding the relative difficulty of an item on the two administrations. First, the difficulties are calculated in somewhat different ways. The 1968 figure is the proportion of students writing the test who chose the correct response to a particular item. The 1976 figure is calculated as follows:

Within each school, the proportion of students writing the test who chose the correct response was calculated. These indices were then averaged over schools to give the corrected difficulty.

Second, as had been mentioned above, there has been considerable change in curriculum over the eight years since the original administration. This cannot help but affect the response pattern to particular items, and it would be unreasonable to draw conclusions about the relative level of achievement of current students without considering this factor.

Third, the two administrations involved somewhat different populations. Those writing the test in 1968 were students who intended to attend university and were writing the test in an effort to increase their chance of acceptance. Students writing in 1976 were doing so whether or not they intended to proceed to university, and without the motivation provided in 1968, since their mark on the test would be of little or no use to them.

Bearing all these caveats in mind, one still finds a striking similarity between the two sets of difficulty indices. The mean difficulty has dropped by only 0.02; the standard deviation has increased by the same amount. Of 36 items, there are only nine on which the absolute shift in difficulty equals or exceeds 0.10; of these nine, four proved substantially easier than in 1968 and five proved substantially harder. The coefficient of correlation between the two sets of figures is 0.84.

2. BISERIAL CORRELATION

The mean biserial correlation coefficient, using uncorrected total test score as the criterion, was 0.46 with a standard deviation of 0.09.

For the 1968 administration, using corrected total test score as the criterion, the mean was 0.46 with a standard deviation of 0.09.

The distribution of biserial correlations for both administrations is given in Table A6.2. Again, caution should be observed in making direct comparisons because of the difference in criteria.

3. RELIABILITY

The Hoyt estimate of reliability for uncorrected scores was 0.79. On the original administration, the estimated reliability coefficient for corrected scores, found by using an adaptation of the KR(20) formula derived by P.L. Dressel, was 0.78.

4. SPEEDEDNESS

A test is considered unspeeded by one rule of thumb if: (a) at least 80 per cent of those writing reach the last item; and (b) all candidates reach the three-quarter mark. On this administration, statistics on the percentage of students not reaching a given item were computed in the same way as were difficulty indices--i.e., by averaging over schools. Detailed figures on item-by-item dropout are given in Table A6.3, along with the same data for the 1968 administration (not computed in this way).

On both administrations the test was somewhat speeded, but less on the present administration. All students did reach the three-quarter mark, with 1 per cent dropout only at item 30. However, only 30 per cent reached the last item (though 86 per cent reached item 34). It should be pointed out that on the last items on a test, the "not reached" figure is subject to

considerable distortion, since it is impossible to distinguish between those who did not reach the items and those who reached them but did not attempt them. In view of the 86 per cent figure at item 34, it seems reasonable to assume that somewhat more than 30 per cent did in fact reach the end, but chose not to attempt the last item or two. There are no grounds for assuming, however, that this figure would approach the 80 per cent requisite for unspeededness.

5. CONTENT VALIDITY

As a check on content validity, teachers of Calculus, of Relations and Functions, and of various first year university Mathematics courses were asked to complete Test Appraisal Inventories related to the test. Secondary teachers were asked to classify each test item under one of the following headings:

A. Old knowledge that students should have on entry to the course.

A1. This knowledge is not reviewed in the course.

A2. This knowledge is reviewed in the course.

B. New knowledge that all students are expected to learn in the course.

C. New knowledge that some students are expected to learn in the course.

C1. Only 1% to 25% of students should learn this.

C2. Only 26% to 50% of students should learn this.

C3. Only 51% to 75% of students should learn this.

C4. More than 75% but not all students should learn this.

D. New knowledge that no student is expected to learn.

The corresponding categories for university teachers were:

A. Old knowledge that students should have on entry to the course.

A1. This knowledge is not reviewed in the course.

A2. This knowledge is reviewed in the course.

B. New knowledge that all students are expected to learn in the course.

C. Other.

Teachers classifying an item other than A1, A2 or B were asked to give a brief explanation of their classification (enrichment material, not relevant to course, etc.). Almost without exception, these comments were statements that the item content was not relevant to a particular course, or that it was enrichment material; therefore, the comments are not reported here in detail.

Table A6.4 presents, for each item, the number of Relations and Functions teachers, the number of Calculus teachers, and the number of first year university Mathematics teachers choosing each classification for the item. The numbers of respondents at each level were: Relations and Functions teachers, 98; Calculus teachers, 74; university teachers, 41.

Table A6.5 presents, for each item, the following information:

- (a) general description of the content. Items judged by the test selection committee to be a review of pre-SSGD work are described as "review".
- (b) assessment given the item by teachers of Relations and Functions, teachers of Calculus, and teachers of first year university Mathematics courses. This assessment was arrived at in the case of secondary teachers by aggregating the number of respondents assessing the item as (in this order) A1, A2, B, C4, C3, C2, C1, and D, until the aggregate number equalled or exceeded 75 per cent of those responding to the item. In the case of university teachers, the aggregation was done in the order A1, A2, B, C, and the cutoff point was 70 per cent to allow for the great diversity in first year university courses. Thus, for example, the rating by Relations and Functions teachers of item 1 as B means that fewer than 75 per cent of these teachers rated it as A1 or A2, but 75 per cent or more rated it A1, A2 or B.
- (c) difficulty of the item (the average over schools of the proportion of students correctly answering the item).

This table provides two criteria for determining content validity. First is the assessment by secondary teachers of whether students should be able to answer a particular item on entry to or on exit from their SSHGD level courses. Of the 36 items, there are eight for which teachers of both courses agree that the student should have this knowledge on entry to SSHGD level courses. There are 25 further items whose content is, according to teachers, taught in at least one of Calculus and Relations and Functions to all students. Of the remaining items, none meets the criterion of 75 per cent classification as A1, A2 or B by teachers of one of these courses. Item 17 is so

classified by 71.1 per cent of Relations and Functions teachers; item 32 is so classified by 61.7 per cent of Calculus teachers; and item 36 is so classified by 60.3 per cent of Calculus teachers.

The second criterion is the assessment by university teachers of whether the student should have met the item content before entry to a first year Mathematics course, or should learn it in that course. Of the 36 items, university teachers judged that a student should know the content of 19 on admission, and should learn the content of a further 12 during his/her first year course (indicating that this content is relevant to first year work in Mathematics). For the remaining five items, the percentages of first year teachers assessing them as A1, A2 or B were respectively 59.0, 51.3, 64.1, 48.8 and 69.3.

Teachers were also asked to comment on the suitability of the test as a whole for assessing the achievement of students in their courses (at the secondary level) or entering their courses (at the university level).

Several secondary teachers were disturbed by the combination of Relations and Functions questions with Calculus questions on the test, although others did not find this a problem. A number of Calculus teachers pointed out areas of omission--integral calculus, limits, polar graphs, length of curves; one of these teachers said that no more than 5 per cent of his/her course was covered on the test. One teacher gave a lengthy list of skills which he/she felt should have been tested, including operations on rational expressions, use of logarithms, ability to discriminate between such expressions as $\sin x$ and $\sin(\underline{x})$, and simplification of complicated expressions (both numerical and algebraic) using factoring. This teacher felt there should be more on trigonometry; one of these felt that the trigonometry that was there was on a Grade Eleven level. One said that the calculus questions in general tested pre-SSHGD knowledge. One teacher described the test as designed "to prove today's Gr. 13 knows no math" and frustrating for the students because the items "are

imaginative, inventive and contain multi-concepts "

Another teacher felt that the existence of alternate methods for doing some items was a problem. The test was described by a few teachers as "designed for the sophisticated mathematician" and "requiring a high level of reasoning and abstraction", thus "testing the ability of the student more than the teacher." A number of teachers described the test as "a good experience" or "fair". One teacher stated, "This material has been taught; the knowledge is there to solve problems if students can tie the ideas together." Two others made similar comments.

The number of comments from university teachers was small. Some of these teachers taught courses for which SSHGD level Mathematics was not a prerequisite (computer science, elementary probability theory, etc.). One teacher found the test dull, claiming there was "not one question that is at all interesting to students at any level whether they can do them or not." Another described it as "Victorian". One felt it contained too much pre-SSHGD material. One felt it contained too much material for a 60-minute test, was repetitious, and should have contained coverage of linear algebra. One felt there was too much emphasis on transformations and mapping notation. One said there was very little on the test that should be "entirely new" at this level, but that students did not in fact have a working knowledge of the material. Two described it as a good test, but one felt that not all entering students knew the content.

In summary, 32 of the 35 items meet the criterion of appropriate assessment by secondary teachers, with a further item very close to that criterion. 31 of the items are assumed by or taught by at least 75 per cent of first year university teachers. Teachers' general comments, although somewhat critical of the actual content of the test (largely as too sophisticated) were more concerned with omissions--largely of many portions of the calculus course.

TABLE A6.1

Mathematics Achievement Test
Corrected Item Difficulties

<u>Item</u>	<u>D for 1968</u> <u>Administration</u>	<u>D for Present</u> <u>Administration</u>	<u>Apparent</u> <u>Shift^a</u>
1	0.82	0.80	-0.02
2	0.59	0.55	-0.04
3	0.83	0.85	+0.02
4	0.40	0.42	+0.02
5	0.56	0.57	+0.01
6	0.76	0.80	+0.04
7	0.81	0.82	+0.01
8	0.24	0.30	+0.06
9	0.72	0.77	+0.05
10	0.88	0.89	+0.01
11	0.36	0.32	-0.04
12	0.40	0.62	+0.22
13	0.91	0.92	+0.01
14	0.19	0.12	-0.07
15	0.67	0.62	-0.05
16	0.56	0.52	-0.04
17	0.65	0.67	+0.02
18	0.58	0.45	-0.13
19	0.61	0.78	+0.17
20	0.51	0.36	-0.15
21	0.70	0.67	-0.03
22	0.54	0.53	-0.01
23	0.31	0.45	+0.14
24	0.76	0.70	-0.06
25	0.27	0.25	-0.02
26	0.33	0.29	-0.04
27	0.36	0.42	+0.06
28	0.59	0.28	-0.31
29	0.39	0.51	+0.12
30	0.36	0.34	-0.02
31	0.57	0.18	-0.39
32	0.61	0.25	-0.36
33	0.51	0.49	-0.02
34	0.40	0.45	+0.05
35	0.19	0.23	+0.04
36	0.07	0.07	0.00

^aSee text for explanation of the different methods used in calculating these figures for the two administrations.

TABLE A6.2

Mathematics Achievement Test
Distribution of Biserial Correlations

<u>Correlation</u>	<u>1968 Administration</u> ^a	<u>Present Administration</u>
0.00 - 0.09	0	0
0.10 - 0.19	0	0
0.20 - 0.29	1	0
0.30 - 0.39	8	7
0.40 - 0.49	14	17
0.50 - 0.59	11	9
0.60 - 0.69	2	3
0.70 - 0.79	0	0
0.80 - 0.89	0	0
0.90 - 0.99	0	0

^aNote that in 1968 the criterion was corrected total test score; on the present administration, the criterion was uncorrected total score.

TABLE A6.3

Mathematics Achievement Test
 Distribution of Students for Whom the Given Item
 Was the First 'Not Reached'

<u>Item</u>	1968 Administration ^a		Present Administration	
	<u>% of Students Not Reaching</u>	<u>Cumula- tive %</u>	<u>% of Students Not Reaching</u>	<u>Cumula- tive %</u>
24	0	0	0	0
25	1	1	0	0
26	0	1	0	0
27	0	1	0	0
28	1	2	0	0
29	0	2	0	0
30	2	4	1	1
31	3	7	1	2
15	0	7	0	2
33	6	13	5	7
34	8	21	7	14
35	25	46	29	43
36	22 ^b	68	27 ^b	70

^aPlease see text for explanation of the different methods used in calculating these figures for the two administrations.

^bFor the last item, the "not reached" figure includes also those students who reached but did not attempt the item.

TABLE A6.4

Mathematics Achievement Test
Classification of Test Items by Teachers

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C(University)</u> <u>D(Secondary)</u>
1	Rel. and F. Calculus University	4 15 28	27 24 5	65 10 0	2 0 -	0 0 -	0 0 -	0 1 -	0 15 6
2	Rel. and F. Calculus University	0 9 22	1 3 6	80 11 1	3 1 -	2 2 -	6 1 -	5 1 -	1 36 10
3	Rel. and F. Calculus University	20 32 21	75 28 13	2 1 0	0 0 -	0 0 -	0 1 -	0 0 -	0 3 5
4	Rel. and F. Calculus University	8 15 21	59 29 10	19 7 1	2 1 -	3 1 -	1 1 -	0 0 -	5 11 7
5	Rel. and F. Calculus University	66 45 35	26 18 3	3 0 0	0 0 -	0 0 -	1 0 -	0 0 -	0 3 1
6	Rel. and F. Calculus University	8 27 13	64 20 16	22 5 0	2 0 -	0 0 -	0 0 -	0 1 -	0 11 10

TABLE A6.4 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C(University) D(Secondary)</u>
7	Rel. and F. Calculus University	12	60	19	2	0	1	0	3
		35	11	6	0	0	0	1	10
		15	20	2	-	-	-	-	4
8	Rel. and F. Calculus University	2	48	42	1	2	0	1	1
		22	22	10	0	0	0	1	10
		13	15	7	-	-	-	-	4
9	Rel. and F. Calculus University	18	12	25	2	0	1	2	31
		3	17	45	2	0	0	1	1
		4	18	9	-	-	-	-	8
10	Rel. and F. Calculus University	0	0	21	0	0	0	0	68
		0	0	69	0	0	0	0	0
		0	19	9	-	-	-	-	11
11	Rel. and F. Calculus University	34	39	6	4	1	3	6	3
		30	21	7	0	0	3	1	4
		24	6	1	-	-	-	-	8
12	Rel. and F. Calculus University	15	28	31	2	3	6	2	8
		17	16	15	1	1	1	6	9
		8	15	6	-	-	-	-	10
13	Rel. and F. Calculus University	1	0	'24	1	0	0	0	61
		0	0	69	0	0	0	0	0
		2	17	9	-	-	-	-	11
14	Rel. and F. Calculus University	71	17	1	0	0	0	0	7
		40	16	1	0	0	0	1	8
		19	4	1	-	-	-	-	15

TABLE A6.4 (continued)

Item	Teachers of	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	C(University) D(Secondary)
15	Rel. and F. Calculus University	3 13 11	13 10 9	74 10 6	3 0 -	2 0 -	2 0 -	0 2 -	1 26 13
16	Rel. and F. Calculus University	20 20 21	16 10 4	37 2 0	1 0 -	3 1 -	5 0 -	5 2 -	9 27 14
17	Rel. and F. Calculus University	4 9 12	8 7 7	57 9 5	3 0 -	2 0 -	7 2 -	8 3 -	8 33 15
18	Rel. and F. Calculus University	37 25 22	33 24 9	18 5 3	2 0 -	1 0 -	3 1 -	0 2 -	3 8 5
19	Rel. and F. Calculus University	35 26 20	41 29 12	14 3 0	0 0 -	2 0 -	1 1 -	2 0 -	1 5 7
20	Rel. and F. Calculus University	2 12 23	1 17 7	88 14 1	2 1 -	1 0 -	2 2 -	2 1 -	0 27 8
21	Rel. and F. Calculus University	49 10 25	26 37 5	10 13 0	0 2 -	0 0 -	0 1 -	1 0 -	9 4 9
22	Rel. and F. Calculus University	48 20 10	41 41 16	1 5 4	1 0 -	1 0 -	1 1 -	0 0 -	3 0 9

TABLE A6.4 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C(University) D(Secondary)</u>
23	Rel. and F. Calculus University	32 14 8	20 29 19	9 9 2	0 0 -	2 1 -	1 0 -	3 0 -	28 13 10
24	Rel. and F. Calculus University	23 2 0	12 6 14	21 60 14	0 1 -	0 0 -	2 0 -	0 0 -	34 0 11
25	Rel. and F. Calculus University	2 1 3	6 3 19	63 53 10	3 0 -	4 4 -	2 2 -	1 2 -	14 3 7
26	Rel. and F. Calculus University	6 22 12	64 25 19	22 9 6	0 0 -	1 0 -	1 0 -	2 1 -	1 8 2
27	Rel. and F. Calculus University	2 15 10	35 29 21	51 10 0	3 0 -	2 0 -	1 0 -	1 0 -	2 9 8
28	Rel. and F. Calculus University	4 21 15	42 18 16	43 10 1	2 1 -	2 0 -	1 1 -	0 1 -	3 12 7
29	Rel. and F. Calculus University	58 35 21	19 19 7	2 1 3	0 0 -	2 0 -	4 2 -	1 0 -	10 9 8
30	Rel. and F. Calculus University	1 6 20	2 10 8	74 11 2	3 1 -	3 0 -	9 2 -	5 3 -	1 28 9

TABLE A6.4 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C(University) D(Secondary)</u>
31	Rel. and F. Calculus University	39 30 21	44 24 10	10 6 3	0 0 -	1 0 -	1 0 -	0 1 -	2 4 5
32	Rel. and F. Calculus University	12 6 3	12 3 11	22 33 15	0 4 -	1 3 -	3 4 -	10 7 -	33 8 10
33	Rel. and F. Calculus University	1 0 0	2 0 17	20 68 12	1 0 -	2 0 -	1 0 -	0 1 -	63 0 10
34	Rel. and F. Calculus University	4 0 1	1 0 12	18 63 14	2 2 -	1 3 -	2 0 -	2 1 -	61 0 12
35	Rel. and F. Calculus University	0 0 1	1 0 8	14 56 19	1 3 -	4 3 -	1 5 -	3 1 -	64 1 11
36	Rel. and F. Calculus University	31 24 10	6 5 8	3 12 10	0 2 -	1 2 -	2 2 -	5 5 -	42 16 11

TABLE A6.5

Mathematics Achievement Test					
Description of Test Items					
Item No.	Content	Assessment by			Difficulty
		R and F Teachers	Calculus Teachers	University Teachers	
1	Conics	B	B	A1	0.80
2	Review (functions)	B	D	A2	0.55
3	Review (trigonometry)	A2	A2	A2	0.85
4	Functions, Trigonometry	B	B	A2	0.42
5	Review (functions)	A2	A2	A1	0.57
6	Phase Shift, Period, Amplitude	A2	B	A2	0.80
7	Function as Mapping	B	B	A2	0.82
8	Inverse of a Function	B	B	A2	0.30
9	Tangents to Curves	D	B	B	0.77
10	Differentiation	D	B	B	0.89
11	Review (functions)	A2	A2	A2	0.32
12	Reflection, Second Degree Relations	B	C4	B	0.62
13	Differentiation	D	B	B	0.92
14	Review (synthetic geometry)	A2	A2	C	0.12
15	Translation	B	D	C	0.62
16	Rotation	B	D	C	0.52
17	Reflection	C3	D	C	0.67
18	Review (functions)	B	A2	A2	0.45
19	Second Degree Relations	A2	A2	A2	0.78
20	Second Degree Relations, Conic's	B	D	A2	0.36
21	Review (algebra)	A2	B	A2	0.67
22	Review (functions)	A2	A2	B	0.53
23	Trigonometry, Review (functions)	D	B	B	0.45
24	Derivatives (velocity, acceleration)	D	B	B	0.70
25	Tangents to Curves	C4	B	B	0.25

TABLE A6.5 (continued)

Item No.	Content	Assessment by			Difficulty
		R and F Teachers	Calculus Teachers	University Teachers	
26	Function as Mapping	B	B	A2	0.29
27	Trigonometry	B	B	A2	0.42
28	Trigonometry	B	B	A2	0.28
29	Review (functions)	A2	A2	A2	0.51
30	Conics	B	D	A2	0.34
31	Review (analytic geometry)	A2	A2	A2	0.18
32	Maxima and Minima	D	C2	B	0.25
33	Rate of Change	D	B	B	0.49
34	Derivatives (velocity, acceleration)	D	B	C	0.45
35	Rate of Change	D	B	B	0.23
36	Area Between Curves	D	C1	B	0.07

APPENDIX A7

TECHNICAL REPORT

TEST OF ARITHMETIC AND BASIC ALGEBRA

The Test of Arithmetic and Basic Algebra was administered to 1,687 students potentially eligible in June 1976 for a Secondary School Graduate Diploma (SSGD). These students were drawn from 52 of the 53 Anglophone schools involved in the study. The remaining school was excluded because it contains only candidates for the Secondary School Honour Graduation Diploma (SSHGD). The population tested included students currently studying Mathematics in the Enriched, Advanced and General streams, as well as students not presently studying Mathematics.

The test consisted of 35 multiple-choice 5-option items providing a broad coverage of basic arithmetic and algebraic skills. The items may be broadly categorized in terms of content as: basic arithmetic, 9 items; basic algebra, 17 items; exponents, 3 items; quadratic equations, 2 items; analytic geometry, 4 items.

A second categorization of the items may be done on the basis of the grade level at which, according to the most recent Ministry of Education guidelines for Mathematics, the student should first have encountered the content of a particular item. This grade level differs for certain items, according to whether a student is enrolled in the General or Advanced stream in Mathematics. Table A7.1 presents for each of these streams the number of items first encountered at each grade level.

1. DIFFICULTY

The difficulty D of an item is generally reported as the proportion of persons writing the test who chose the correct response to the item. In the case of this test, what is reported is an average difficulty for each item. The difficulty was calculated in the usual way for the students within each individual school. The resulting indices were then averaged over all schools.

The mean item difficulty was 0.56 with a standard deviation of 0.16. The distribution of item difficulties is given in Table A7.2.

2. BISERIAL CORRELATION

The mean biserial correlation, using raw total test score as the criterion, was 0.63, with a standard deviation of 0.09. The distribution of biserial correlations is given in Table A7.3.

3. RELIABILITY

The Hoyt estimate of reliability for uncorrected scores was 0.91.

4. SPEEDEDNESS

A test is considered unspeeded by one rule of thumb if: (a) at least 80 per cent of those writing reach the last item; and (b) all candidates reach the three-quarter mark. 80 per cent of the candidates on this test attempted the last item, and 99 per cent

reached the three-quarter mark. The test was thus virtually unspeeded. An item-by-item report on dropouts is given in Table A7.4.

5. TEST SCORES

A correction for guessing was applied to test scores by subtracting from the number of correct responses one-quarter mark for each incorrect response.

The mean corrected test score was 16.58, with a standard deviation of 9.39.

6. CONTENT VALIDITY

As a check on content validity, teachers of Applications of Mathematics 2 (the General stream SSGD level course), Foundations of Mathematics 2 (the Advanced stream SSGD level course), and various first year CAAT Mathematics courses completed Test Appraisal Inventories related to the test. The secondary teachers were asked to classify each item under one of the following headings:

A. Old knowledge that students should have on entry to the course:

A1. This knowledge is not reviewed in the course.

A2. This knowledge is reviewed in the course.

B. New knowledge that all students are expected to learn in the course.

C. New knowledge that some students are expected to learn in the course.

C1. Only 1% to 25% of students should learn this.

C2. Only 26% to 50% of students should learn this.

C3. Only 51% to 75% of students should learn this.

C4. More than 75% but not all students should learn this.

D. New knowledge that no student is expected to learn.

The corresponding categories for CAAT teachers were:

A. Old knowledge that students should have on entry to the course:

A1. This knowledge is not reviewed in the course.

A2. This knowledge is reviewed in the course.

B. New knowledge that all students are expected to learn in the course.

C. Other.

Table A7.5 presents, for each item, the number of teachers at each level choosing each classification for the item. The numbers of respondents at each level were: Applications teachers, 77; Foundations teachers, 93; CAAT teachers, 37.

Teachers classifying an item other than A1, A2 or B were asked to give a brief explanation of their classification (enrichment material, not relevant to course, etc.). These explanations, where they occurred, were almost without exception statements that the material was not relevant to a particular

course (for example, 126 of 127 comments by CAAT teachers were of this nature). The comments are therefore not reported here in detail. Table A7.6 presents, for each item on the test, the following information:

- (a) general description of the content.
- (b) statement of the grade level at which students in the General or Advanced stream would first be expected to encounter the item content, according to Ministry of Education guidelines.
- (c) assessment given the item by teachers of the Applications course, teachers of the Foundations course, and teachers of first year CAAT Mathematics courses. This assessment was arrived at by aggregating the number of respondents assessing the item as (in this order) A1, A2, B, C4, C3, C2, C1, and D, until the aggregate number equalled or exceeded 75 per cent of those responding to the item. (In the case of CAAT teachers, the aggregation was done in the order A1, A2, B, C.) Thus, for example, the rating by Applications teachers of item 3 as A2 means that fewer than 75 per cent of these teachers rated the item as A1, but 75 per cent or more of them rated it as A1 or A2.
- (d) difficulty of the item (the average over schools of the proportion of students correctly answering the item).

This table includes three different criteria for determining content validity. First is the level at which, according to Ministry guidelines, the student should first have been taught the content of an item. All test items meet this criterion, since their content is included in the guidelines either at or before the SSGD level. Second is the assessment by teachers of Applications and Foundations as to whether students should be able to answer a particular item on entry to or on exit from their

courses. Again the content of all items was judged to fall either at a pre-SSGD level (25 items in the General stream, 30 items in the Advanced stream) or at the SSGD level (all remaining items). Third is the assessment by CAAT teachers of their expectations of entering students and the content of their first year courses. These teachers expected the students to have encountered the content of 15 of the items before entering college, and taught the content of a further 17 (indicating that the content of these items is relevant to first year work in Mathematics). The remaining three items did not meet the criterion of 75 per cent of the CAAT teachers expecting or teaching the content, but came close, with the aggregated totals at the B level being respectively 68 per cent, 65 per cent and 73 per cent.

Teachers at both levels were also asked to make general comments on the suitability of the test as a whole. Ten secondary teachers stated that the test was not a good measure of SSGD level mathematical skills, since the content was almost all at a lower level. (This, of course, was to be expected, since the test was not designed as a measure of competence in SSGD level content, but as a test of basic skills.) Three other secondary teachers stated, on the other hand, that some or all of the test was too difficult for SSGD level students in the General stream, or for students not currently taking Mathematics. Two teachers stated that they thought the test was suitable--a "good sample for basic applications", "should be of average difficulty". One teacher said the test was not an accurate measure of achievement, but did not specify why.

Eight CAAT teachers described the test as suitable in that all entering students should be able to handle all the test content. However, five of these teachers and seven other teachers said that in fact their incoming students do not have these skills. Three CAAT teachers deplored the omission of a question on ratio and proportion; two suggested that word problems should have been included; and one made a general statement that the

test was "too limited to evaluate the broad range of items that the students should bring with them."

In general, then, the test appears to be a reasonably valid instrument in terms of content. The majority of criticisms at the secondary level were statements that the test was not what it was not intended to be--a measure of achievement skills taught at the SSGD level. CAAT teachers generally approved of the content, while pointing out two areas omitted which they felt to be of importance.

TABLE A7.1

Test of Arithmetic and Basic Algebra
 Number of Test Items Whose Content is First Encountered at
 Various Grade Levels According to Ministry
 of Education Guidelines

<u>Grade Level</u>	<u>Number of Items</u>	
7	7	
8	1	
	General stream	Advanced stream
9	10	16
10	6	4
11	7	5
12	4	2

TABLE A7.2

Test of Arithmetic and Basic Algebra
Distribution of Average Item Difficulties

<u>Difficulty</u>	<u>Number of items</u>
0.00 - 0.09	0
0.10 - 0.19	0
0.20 - 0.29	1
0.30 - 0.39	6
0.40 - 0.49	6
0.50 - 0.59	7
0.60 - 0.69	7
0.70 - 0.79	7
0.80 - 0.89	1
0.90 - 0.99	0

TABLE A7.3

Test of Arithmetic and Basic Algebra
Distribution of Biserial Correlations

<u>Correlation</u>	<u>Number of items</u>
0.00 - 0.09	0
0.10 - 0.19	0
0.20 - 0.29	0
0.30 - 0.39	0
0.40 - 0.49	0
0.50 - 0.59	14
0.60 - 0.69	13
0.70 - 0.79	7
0.80 - 0.89	1
0.90 - 0.99	0

TABLE A7.4
Test of Arithmetic and Basic Algebra
 Distribution of Students for Whom the Given Item
 Was the First "Not Reached"

<u>Item</u>	<u>% of Students Not Reaching</u>	<u>Cumulative %</u>
25	1	1
26	0	1
27	0	1
28	0	1
29	0	1
30	0	1
31	1	2
32	0	2
33	2	4
34	1	5
35	15 ^a	20

^aFor the last item, the "not reached" figure includes also those students who reached but did not attempt the item.

TABLE A7.5

Test of Arithmetic and Basic Algebra
Classification of Test Items by Teachers

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C (CAAT)</u> <u>D(Secondary)</u>
1	Applications Foundations CAAT	59 86 17	17 7 20	0 0 0	0 0 -	0 0 -	0 0 -	0 0 -	0 0 0
2	Applications Foundations CAAT	69 86 26	7 5 11	0 0 0	1 0 -	0 0 -	0 0 -	0 1 -	0 1 0
3	Applications Foundations CAAT	54 84 21	23 9 13	0 0 1	0 0 -	0 0 -	0 0 -	0 0 -	0 0 2
4	Applications Foundations CAAT	53 80 20	24 13 17	0 0 0	0 0 -	0 0 -	0 0 -	0 0 -	0 0 0
5	Applications Foundations CAAT	37 64 14	39 29 23	0 0 0	0 0 -	0 0 -	0 0 -	0 0 -	0 0 0
6	Applications Foundations CAAT	55 85 24	21 8 12	0 0 1	0 0 -	0 0 -	0 0 -	0 0 -	0 0 0

TABLE A7.5 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C (CAAT)</u> <u>D(Secondary)</u>
7	Applications	57	18	0	0	0	0	0	0
	Foundations	81	12	0	0	0	0	0	0
	CAAT	24	13	0	-	-	-	-	0
8	Applications	26	40	8	0	1	0	0	0
	Foundations	33	50	8	1	1	0	0	0
	CAAT	5	24	5	-	-	-	-	3
9	Applications	33	35	4	0	0	2	0	2
	Foundations	69	19	2	0	0	1	0	2
	CAAT	14	18	3	-	-	-	-	2
10	Applications	9	36	20	5	1	1	0	3
	Foundations	11	43	36	3	0	0	0	0
	CAAT	2	24	7	-	-	-	-	4
11	Applications	11	41	16	4	2	0	1	1
	Foundations	7	48	36	2	0	0	0	0
	CAAT	0	27	6	-	-	-	-	4
12	Applications	8	25	29	2	3	4	1	3
	Foundations	5	30	54	0	4	0	0	0
	CAAT	0	26	7	-	-	-	-	4
13	Applications	31	39	4	0	1	0	0	1
	Foundations	60	32	1	0	0	0	0	0
	CAAT	6	25	4	-	-	-	-	2

TABLE A7.5 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C</u> <u>D</u> (CAAT) (Secondary)
14	Applications Foundations CAAT	24 54 6	40 35 25	10 4 4	0 0 -	0 0 -	1 0 -	1 0 -	0 0 2
15	Applications Foundations CAAT	30 81 7	42 12 26	2 0 0	1 0 -	1 0 -	0 0 -	0 0 -	0 0 4
16	Applications Foundations CAAT	23 73 6	48 20 22	4 0 5	1 0 -	0 0 -	0 0 -	0 0 -	0 0 4
17	Applications Foundations CAAT	15 63 3	53 30 23	5 0 6	0 0 -	1 0 -	2 0 -	0 0 -	0 0 5
18	Applications Foundations CAAT	18 43 1	49 47 25	7 2 6	1 1 -	0 0 -	1 0 -	0 0 -	0 0 5
19	Applications Foundations CAAT	8 39 2	55 52 23	11 2 6	0 0 -	1 0 -	0 0 -	1 0 -	0 0 6
20	Applications Foundations CAAT	9 32 2	45 59 22	18 2 6	1 0 -	2 0 -	0 0 -	0 0 -	1 0 7

TABLE A7.5 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C</u> <u>D</u> (Secondary) (CAAT)
21	Applications Foundations CAAT	13 46 2	54 47 22	7 0 6	0 0 -	2 0 -	0 0 -	0 0 -	0 0 7
22	Applications Foundations CAAT	8 23 2	44 65 24	22 5 4	0 0 -	1 0 -	0 0 -	1 0 -	0 0 7
23	Applications Foundations CAAT	25 68 4	47 23 26	3 2 5	0 0 -	0 0 -	0 0 -	0 0 -	1 0 2
24	Applications Foundations CAAT	20 29 0	46 60 25	7 4 6	0 0 -	1 0 -	0 0 -	0 0 -	1 0 6
25	Applications Foundations CAAT	2 5 0	12 41 21	50 44 8	0 1 -	4 2 -	1 0 -	3 0 -	4 0 8
26	Applications Foundations CAAT	1 1 0	5 6 20	57 82 9	4 2 -	1 1 -	2 1 -	2 0 -	4 0 8
27	Applications Foundations CAAT	31 36 1	34 51 20	6 6 11	0 0 -	1 0 -	0 0 -	1 0 -	2 0 5

TABLE A7.5 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C (CAAT) D (Secondary)</u>
28	Applications Foundations CAAT	29 28 1	32 60 13	8 5 11	0 0 -	0 0 -	2 0 -	0 0 -	3 0 12
29	Applications Foundations CAAT	29 28 1	36 61 11	4 4 16	0 0 -	1 0 -	1 0 -	0 0 -	4 0 9
30	Applications Foundations CAAT	28 31 1	30 56 11	4 6 12	0 0 -	2 0 -	1 0 -	1 0 -	7 0 13
31	Applications Foundations CAAT	17 44 1	43 47 28	11 2 3	1 0 -	0 0 -	0 0 -	1 0 -	1 0 5
32	Applications Foundations CAAT	14 42 1	45 47 19	12 4 8	1 0 -	1 0 -	1 0 -	0 0 -	2 0 9
33	Applications Foundations CAAT	13 33 1	42 55 19	14 5 7	1 0 -	0 0 -	3 0 -	0 0 -	3 0 10
34	Applications Foundations CAAT	9 30 1	29 58 21	26 4 8	1 0 -	1 0 -	4 0 -	2 0 -	4 1 7

TABLE A7.5 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C (CAAT) D(Secondary)</u>
35	Applications	13	28	16	0	1	6	3	7
	Foundations	27	60	3	0	0	0	1	2
	CAAT	1	21	7	-	-	-	-	7

TABLE A7.6

Test of Arithmetic and Basic Algebra
Description of Test Items

Item No.	Content	First Occurrence in Guideline for		Assessment By			Difficulty
		General Students	Advanced Students	Applications Teachers	Foundations Teachers	CAAT Teachers	
1	Basic arithmetic	7		A1	A1	A2	0.71
2	Basic arithmetic	7		A1	A1	A2	0.42
3	Basic arithmetic	8		A2	A1	A2	0.55
4	Basic arithmetic	7		A2	A1	A2	0.70
5	Basic arithmetic	9	9	A2	A2	A2	0.70
6	Basic arithmetic	7		A2	A1	A2	0.72
7	Basic arithmetic	7		A1	A1	A2	0.76
8	Basic arithmetic	7		A2	A2	A2	0.38
9	Basic arithmetic	7		A2	A2	A2	0.31
10	Exponents	10	9	B	B	B	0.37
11	Exponents	10	9	B	B	B	0.53
12	Exponents	11	12	B	B	B	0.44
13	Basic algebra	9	9	A2	A2	A2	0.68
14	Basic algebra	9	9	A2	A2	A2	0.72
15	Basic algebra	9	9	A2	A1	A2	0.58
16	Basic algebra	9	9	A2	A2	A2	0.88
17	Basic algebra	11	11	A2	A2	B	0.68
18	Basic algebra	10	9	A2	A2	B	0.63
19	Basic algebra	10	9	A2	A2	B	0.50
20	Basic algebra	9	9	B	A2	B	0.68

TABLE A7.6 (continued)

Item No.	Content	First Occurrence in Guideline for		Assessment By		
		General Students	Advanced Students	Applications Teachers	Foundations Teachers	CAAT Teachers
21	Basic algebra	9	9	A2	A2	B
22	Basic algebra	9	9	B	A2	B
23	Basic algebra	9	9	A2	A2	A2
24	Basic algebra	10	10	A2	A2	B
25	Quadratic equations	12	11	B	B	B
26	Quadratic equations	12	12	B	B	B
27	Analytic geometry	10	9	A2	A2	B
28	Analytic geometry	11	10	A2	A2	C
29	Analytic geometry	11	10	A2	A2	B
30	Analytic geometry	11	10	A2	A2	C
31	Basic algebra	9	9	A2	A2	A2
32	Basic algebra	11	11	A2	A2	B
33	Basic algebra	11	11	B	A2	C
34	Basic algebra	12	9	B	A2	B
35	Basic algebra	12	11	B	A2	B

0.79
0.69
0.57
0.57
0.48
0.25
0.39
0.66
0.42
0.46
0.41
0.65
0.31
0.33
0.52

APPENDIX A8

TECHNICAL REPORT ON THE

TEST DE RENDEMENT EN MATHÉMATIQUES

The Test de rendement en mathématiques was administered to 140 students potentially eligible in June 1976 for a Secondary School Honour Graduation Diploma (SSHGD). These students were drawn from the Francophone population of the 14 Francophone and bilingual schools involved in the study (for a definition of the Francophone population in bilingual schools, see Chapter two of the report, Part A, subsection 1.1). The population tested included students enrolled during the school year 1976-1976, in both Calcul and Relations et fonctions.

1. TEST CONTENT

The content of the test was identical, except for language, with that of the Mathematics Achievement Test. For a description of the content, see the Technical Report on that test (Appendix A6).

2. TECHNICAL ISSUES

2.1 Difficulty

The difficulty D of an item is generally reported as the proportion of persons writing the test who chose the correct

response to the item. In the case of this test, what is reported is an average difficulty for each item. The difficulty was calculated in the usual way for the students within each individual school. The resulting indices were then averaged over all schools.

The mean item difficulty was 0.44 with a standard deviation of 0.25. The distribution of item difficulties is given in Table A8.1.

2.2 Biserial Correlation

The mean biserial correlation, using uncorrected total test score as the criterion, was 0.49 with a standard deviation of 0.16. The distribution of biserial correlations is given in Table A8.2.

2.3 Reliability

The Hoyt estimate of reliability for uncorrected scores was 0.79.

2.4 Speededness

A test is considered unspeeded by one rule of thumb, if: (a) at least 80 per cent of those writing reach the last item; and (b) all candidates reach the three-quarter mark. This test was somewhat speeded. Although 98 per cent of candidates reached the three-quarter mark, only 27 per cent reached the last item. It should be pointed out that on the last items on a test, the "not reached" figure is subject to considerable distortion, since it is impossible to distinguish between those who did not reach the items and those who reached them but did not attempt them. In view of the 84 per cent figure at item 34 (of 36 items), it seems reasonable to assume that somewhat more than 27 per cent did in fact reach the end, but chose not to attempt the last item

or two. There are no grounds for assuming, however, that this figure would approach the 80 per cent requisite for unspeededness.

Table A8.3 presents an item-by-item account of dropouts.

2.5 Test Scores

A correction for guessing was applied to test scores by subtracting from the number of correct responses one-quarter mark for each incorrect response.

3. TEST APPRAISAL

As a check on content validity, teachers of Calcul, of Relations et fonctions, and of various first year university mathematics courses were asked to complete Test Appraisal Inventories (Test-inventaires estimatifs) related to the test. Secondary teachers were asked to classify each test item under one of the following headings (see the Technical Report on the Mathematics Achievement Test, Appendix A6, for the English translation):

A. Connaissances antérieures que les étudiants devraient avoir au début du cours.

A1. Ces connaissances ne sont pas revues dans le cours.

A2. Ces connaissances sont revues dans le cours.

B. Connaissances nouvelles que tous les étudiants doivent acquérir pendant le cours.

C. Connaissances nouvelles que quelques étudiants doivent acquérir pendant le cours.

C1. De 1% à 25% des étudiants seulement doivent les acquérir.

C2. De 26% à 51% des étudiants seulement doivent les acquérir.

C3. De 51% à 75% des étudiants seulement doivent les acquérir.

C4. Plus de 75% des étudiants--mais pas tous--doivent les acquérir.

D. Connaissances nouvelles qu'aucun étudiant ne doit acquérir pendant le cours.

The corresponding categories for university teachers were:

A. Connaissances antérieures que les étudiants devraient avoir au début du cours.

A1. Ces connaissances ne sont pas revues dans le cours.

A2. Ces connaissances sont revues dans le cours.

B. Connaissances nouvelles que tous les étudiants doivent acquérir pendant le cours.

C. Autre.

Table A8.4 presents, for each item, the number of teachers of Relations et fonctions and the number of teachers of Calcul choosing each classification for the item. The number of respondents was: teachers of Relations et fonctions, 12; teachers of Calcul, 11. There was only one university respondent. Clearly

no conclusions can be drawn from one response, but as a matter of interest this teacher's responses are indicated in the table by an asterisk.

Table A8.5 presents, for each item, the following information:

- (a) A general description of the content. Items judged by the test selection committee to be a review of pre-SSHGD work are described as "review".
- (b) The assessment given the item by teachers of Relations et fonctions and by teachers of Calcul. This assessment was arrived at by aggregating the number of respondents assessing the item as (in this order) A1, A2, B, C4, C3, C2, C1, and D, until the aggregate number equalled or exceeded 75 per cent of those responding to the item. Thus, for example, the rating by teachers of Relations et fonctions of item 3 as A2 means that fewer than 75 per cent of these teachers rated it as A1, but 75 per cent or more rated it as A1 or A2. (Because of the small number of respondents, this method of assigning an overall assessment is imprecise, and should be considered only a rough approximation.)
- (c) The difficulty of the item (the average over schools of the proportion of students correctly answering the item).

This table provides a criterion (admittedly rough) for determining content validity. Of the 36 items, there are five for which teachers of both courses agree that the student should have the necessary knowledge on entry to SSHGD level courses. The content of one item is expected on entry by teachers of Relations and fonctions, although classified C2 by teachers of Calcul. There are 25 further items whose content is, according to teachers, taught in at least one of Calcul et Relations et fonctions to all students. Of the five remaining items, none

meets the criterion of 75 per cent classification as A1, A2 or B by teachers of one of these courses. Item 17 was so classified by 66.7 per cent of Relations et fonctions teachers; item 21 by 63.7 per cent of Calcul teachers; item 23 by 66.6 per cent of Relations et fonctions teachers and 63.7 per cent of Calcul teachers; item 32 by 63.6 per cent of Calcul teachers; and item 36 by 54.6 per cent of Calcul teachers.

Thus 31 of the 36 items meet the criterion of appropriate assessment by secondary teachers, with four of the remaining five close to that criterion. (Considering the number of respondents, a difference of 10 per cent represents about one response.)

TABLE A8.1

Test de rendement en mathématiques
Distribution of Average Item Difficulties

<u>Difficulty</u>	<u>Number of Items</u>
0.00 - 0.09	3
0.10 - 0.19	4
0.20 - 0.29	6
0.30 - 0.39	4
0.40 - 0.49	4
0.50 - 0.59	4
0.60 - 0.69	6
0.70 - 0.79	1
0.80 - 0.89	2
0.90 - 0.99	2

TABLE A8.2

Test de rendement en mathématiques
Distribution of Biserial Correlations

<u>Correlation</u>	<u>Number of Items</u>
0.00 - 0.09	0
0.10 - 0.19	0
0.20 - 0.29	5
0.30 - 0.39	6
0.40 - 0.49	9
0.50 - 0.59	8
0.60 - 0.69	5
0.70 - 0.79	2
0.80 - 0.89	1
0.90 - 0.99	0

TABLE A8.3

Test de rendement en mathématiques
 Distribution of Students for Whom the Given Item
 Was the First "Not Reached"

<u>Item</u>	<u>% of Students Not Reaching</u>	<u>Cumulative %</u>
26	0	0
27	2	2
28	0	2
29	0	2
30	1	3 "
31	0	3
32	0	3
33	7	10
34	6	16
35	32	48
36	25	73

TABLE A8.4

Test de rendement en mathématiques
Classification of Test Items by Teachers

Item	Teachers of	A1	A2	B	C4	C3	C2	C1	C(University) D(Secondary)
1	Rel. et f. Calcul University	0 3 a	4 4	8 2	0 0	0 0	0 0	0 0	0 2
2	Rel. et f. Calcul University	0 3 a	0 1	11 3	1 0	0 0	0 0	0 0	0 4
3	Rel. et f. Calcul University	2 4	9 7 a	1 0	0 0	0 0	0 0	0 0	0 0
4	Rel. et f. Calcul University	1 5	3 3 a	8 2	0 0	0 1	0 0	0 0	0 0
5	Rel. et f. Calcul University	8 4 a	4 5	0 1	0 0	0 0	0 0	0 1	0 0
6	Rel. et f. Calcul University	0 5	5 4	5 1 a	1 0	1 0	0 0	0 1	0 0

TABLE A8.4 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C(University) D(Secondary)</u>
7	Rel. et f. Calcul University	2 6	9 3 a	1 0	0 0	0 0	0 0	0 0	0 2
8	Rel. et f. Calcul University	2 3	6 5	4 1 a	0 0	0 0	0 0	0 0	0 2
9	Rel. et f. Calcul University	4 0	1 4	0 6 a	1 1	0 0	0 0	0 0	5 0
10	Rel. et f. Calcul University	0 0	1 1	1 9 a	1 1	0 0	0 0	0 0	8 0
11	Rel. et f. Calcul University	6 4	2 2 a	2 2	0 1	0 0	0 0	1 0	1 2
12	Rel. et f. Calcul University	1 5	5 2 a	4 1	0 1	1 0	0 0	0 0	1 2
13	Rel. et f. Calcul University	0 0	2 0	1 10 a	1 1	0 0	0 0	0 0	7 0
14	Rel. et f. Calcul University	10 6	0 2	0 1	0 0	0 0	0 0	0 0	1 2 a

TABLE A8.4 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C(University) D(Secondary)</u>
15	Rel. et f. Calcul University	0 3	1 1 a	11 2	0 0	0 0	0 0	0 0	0 5
16	Rel. et f. Calcul University	0 2	1 0	9 0	0 0	0 0	1 0	0 1	1 7 a
17	Rel. et f. Calcul University	0 3	0 0	8 1	1 1	0 0	1 0	0 0	2 6 a
18	Rel. et f. Calcul University	3 1	5 7 a	1 1	0 1	0 0	1 0	0 0	2 1
19	Rel. et f. Calcul University	4 3 a	4 4	2 1	0 1	1 0	0 0	0 0	1 2
20	Rel. et f. Calcul University	0 2	0 1 a	11 3	1 0	0 0	0 0	0 0	0 5
21	Rel. et f. Calcul University	3 1	0 4	1 2	0 0	0 2	0 1	0 0	7 1 a
22	Rel. et f. Calcul University	5 4	7 7 a	0 0	0 0	0 0	0 0	0 0	0 0

TABLE A8.4 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C(University) D(Secondary)</u>
23	Rel. et f. Calcul University	4 3	4 3 a	0 1	0 1	0 0	2 0	0 0	2 3
24	Rel. et f. Calcul University	3 0	2 0 a	2 10	0 0	0 0	1 1	0 0	4 0
25	Rel. et f. Calcul University	1 0	0 1	8 7 a	1 1	1 0	0 0	0 1	1 1
26	Rel. et f. Calcul University	3 3	5 7 a	4 0	0 0	0 0	0 0	0 1	0 0
27	Rel. et f. Calcul University	0 4	4 3 a	8 3	0 0	0 0	0 0	0 1	0 0
28	Rel. et f. Calcul University	0 2	4 5 a	6 1	1 0	0 0	0 0	1 0	0 3
29	Rel. et f. Calcul University	9 6 a	2 2	0 0	0 0	0 0	0 1	1 1	0 1
30	Rel. et f. Calcul University	0 1 a	0 1	9 3	0 1	0 0	1 0	1 0	1 5

TABLE A8.4 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C(University) D(Secondary)</u>
31	Rel. et f. Calcul University	8 6 a	3 3	1 1	0 0	0 0	0 0	0 0	0 1
32	Rel. et f. Calcul University	2 1	3 1	2 5 a	1 1	0 0	1 1	0 1	3 1
33	Rel. et f. Calcul University	1 0	1 1	2 9 a	1 1	0 0	0 0	0 0	6 0
34	Rel. et f. Calcul University	1 0	1 0	0 10 a	0 0	0 0	2 1	0 0	7 0
35	Rel. et f. Calcul University	1 0	1 0	0 9 a	1 0	0 0	0 1	1 0	7 1
36	Rel. et f. Calcul University	3 2	0 1	0 3	0 2	0 0	0 0	1 0	7 3 a

^aIndicates assessment made by sole university respondent.

TABLE A8.5

Test de rendement en mathématiques
Description of Test Items

Item No.	Content	Assessment by		Difficulty
		Teachers of R. et F.	Teachers of Calcul	
1	Conics	B	B	0.76
2	Review (functions)	B	D	0.39
3	Review (trigonometry)	A2	A2	0.80
4	Functions, Trigonometry	B	B	0.25
5	Review (functions)	A2	A2	0.43
6	Phase Shift, Period, Amplitude	B	A2	0.69
7	Function as Mapping	A2	A2	0.91
8	Inverse of a Function	B	B	0.24
9	Tangents to Curves	D	B	0.63
10	Differentiation	D	B	0.89
11	Review (functions)	B	C4	0.18
12	Reflection, Second Degree Relations	B	C4	0.58
13	Differentiation	D	B	0.90
14	Review (synthetic geometry)	A1	B	0.09
15	Translation	B	D	0.65
16	Rotation	B	D	0.50
17	Reflection	C4	D	0.60
18	Review (functions)	B	B	0.60
19	Second Degree Relations	B	C4	0.59
20	Second Degree Relations, Conics	B	D	0.27
21	Review (algebra)	D	C3	0.58
22	Review (functions)	A2	A2	0.42
23	Trigonometry, Review (functions)	C2	D	0.33
24	Derivatives (velocity, acceleration)	D	B	0.68
25	Tangents to Curves	B	C4	0.21

TABLE A8.5 (continued)

Item No.	Content	Assessment by Teachers		Difficulty
		of R. et F.	of Calcul	
26	Function as Mapping	B	A2	0.23
27	Trigonometry	B	B	0.32
28	Trigonometry	B	D	0.07
29	Review (functions)	A2	C2	0.45
30	Conics	B	D	0.28
31	Review (analytic geometry)	A2	A2	0.10
32	Maxima and Minima	C2	C2	0.17
33	Rate of Change	D	B	0.31
34	Derivatives (velocity, acceleration)	D	B	0.42
35	Rate of Change	D	B	0.16
36	Area Between Curves	D	D	0.03

APPENDIX A9

TECHNICAL REPORT ON THE

TEST D'ARITHMETIQUE ET D'ALGÈBRE DE BASE

The Test d'arithmétique et d'algèbre de base was administered to 291 students potentially eligible in June 1976 for a Secondary School Graduation Diploma (SSGD). These students were drawn from the Francophone population of the 14 Francophone and bilingual schools involved in the study (see body of report for definition of Francophone population in bilingual schools). The population tested included students currently studying Mathematics at the Enriched, Advanced and General levels, as well as students not currently studying mathematics.

1. TEST CONTENT

The test content was identical, except for language, with that of the Test of Arithmetic and Basic Algebra. For a description of the content, see the technical report on that test (Appendix A7).

2. TECHNICAL ISSUES ABOUT THE TEST

2.1 Difficulty

The difficulty D of an item is generally reported as the proportion of persons writing the test who chose the correct

response to the item. In the case of this test, what is reported is an average difficulty for each item. The difficulty was calculated in the usual way for the students within each individual school. The resulting indices were then averaged over all schools. The mean item difficulty was 0.42 with a standard deviation of 0.15. The distribution of item difficulties is given in Table A9.1.

2.2 Biserial Correlation

The mean biserial correlation, using raw total test score as the criterion, was 0.66 with a standard deviation of 0.13. The distribution of biserial correlations is given in Table A9.2.

2.3 Reliability

The Hoyt estimate of reliability for uncorrected scores was 0.91.

2.4 Speededness

A test is considered unspeeded by one rule of thumb, if: (i) at least 80 per cent of those writing reach the last item; (ii) all candidates reach the three-quarter mark. This test evinces some speededness. Only 74 per cent of the candidates reached the last item, and only 95 per cent reached the three-quarter mark. It should be noted that the "not reached" figure for the last item is subject to considerable distortion, since it is impossible to distinguish between those who did not reach the item and those who reached it but did not attempt it. It is possible, therefore, in view of the fact that 90 per cent of those writing reached the next-to-last item, that in fact 80 per cent or more did reach this item but a number of those did not attempt it. An item-by-item report on dropouts is given in Table A9.3.

2.5 Test Scores

A correction for guessing was applied to test scores by subtracting from the number of correct responses one-quarter mark for each incorrect response.

The mean corrected test score was 10.67 with a standard deviation of 9.68.

3. TEST APPRAISAL

3.1 Content Validity

As a check on content validity, teachers of Applications des Mathématiques 2 (the General stream SSGD level course), Fondement des Mathématiques 2 (the Advanced stream SSGD level course), and various first year CAAT mathematics courses taught in French completed Test Appraisal Inventories (Test inventaires estimatifs) related to the test. The secondary teachers were asked to classify each item under one of the following headings (see Technical Report on the Test of Arithmetic and Basic Algebra for the English translation):

A. Connaissances antérieures que les étudiants devraient avoir au début du cours.

A1. Ces connaissances ne sont pas revues dans le cours.

A2. Ces connaissances sont revues dans le cours.

B. Connaissances nouvelles que tous les étudiants doivent acquérir pendant le cours.

C. Connaissances nouvelles que quelques étudiants doivent acquérir pendant le cours.

C1. De 1% à 25% des étudiants seulement doivent les acquérir.

C2. De 26% à 50% des étudiants seulement doivent les acquérir.

C3. De 51% à 75% des étudiants seulement doivent les acquérir.

C4. Plus de 75% des étudiants--mais pas tous--doivent les acquérir.

D. Connaissances nouvelles qu'aucun étudiant ne doit acquérir pendant le cours.

The corresponding categories for CAAT teachers were:

A. Connaissances antérieures que les étudiants devraient avoir au début du cours.

A1. Ces connaissances ne sont pas revues dans le cours.

A2. Ces connaissances sont revues dans le cours.

B. Connaissances nouvelles que tous les étudiants doivent acquérir pendant le cours.

C. Autre.

Table A9.4 presents, for each item, the number of teachers at each level choosing each classification for the item. The numbers of respondents at each level were: Applications teachers, 15; Fondement teachers, 12; CAAT teachers, 19. Table A9.5 presents, for each item on the test, the following information:

(a) A general description of the content.

- (b) A statement of the grade level at which students in the General or Advanced stream would first be expected to encounter the item content, according to Ministry of Education guidelines.
- (c) The assessment given the item by teachers of the Applications course, teachers of the Fondement course, and teachers of first year CAAT Mathematics courses. This assessment was arrived at by aggregating the number of respondents assessing the item as (in this order) A1, A2, B, C4, C3, C2, C1, and D, until the aggregate number equalled or exceeded 75 per cent of those responding to the item. (In the case of CAAT teachers the aggregation was done in the order A1, A2, B, C.) Thus for example the rating by Applications teachers of item 1 as A2 means that fewer than 75 per cent of these teachers rated the item as A1, but 75 per cent or more of them rated it as A1 or A2.
- (d) The difficulty of the item (the average over schools of the proportion of students correctly answering the item).

This table includes three different criteria for determining content validity. First is the level at which, according to Ministry guidelines, the student should first have been taught the content of an item. All test items meet this criterion, since their content is included in the guidelines either at or before the SSGD level. Second is the assessment by teachers of Applications and Fondement as to whether students should be able to answer a particular item on entry to or on exit from their courses. All items meet this criterion for the Advanced level, where students are expected to have met the content of all items either on entry to the SSGD level course (27 items) or during that course (8 items). At the General level, students are expected to have met the content of 25 items before the SSGD level, and the content of a further 9 items at that level. The one remaining item does not quite meet the criterion; slightly

fewer than 75 per cent of Applications teachers rated it as A1, A2, or B, but 80 per cent rate it as A1, A2, B or C. The third criterion is the assessment by CAAT teachers of their expectations of entering students and the content of their first year courses. These teachers expected the students to have encountered the content of 24 of the items before entering college, and taught the content of the remaining 11 (indicating that the content of these items is relevant to first year work in mathematics).

It thus appears that the test can be considered valid in terms of content.

TABLE A9.1

Test d'arithmétique et d'algèbre de base
Distribution of Average Item Difficulties

<u>Difficulty</u>	<u>Number of Items</u>
0.00 - 0.09	0
0.10 - 0.19	2
0.20 - 0.29	8
0.30 - 0.39	5
0.40 - 0.49	9
0.50 - 0.59	8
0.60 - 0.69	2
0.70 - 0.79	1
0.80 - 0.89	0
0.90 - 0.99	0

TABLE A9.2

Test d'arithmétique et d'algèbre de base
Distribution of Biserial Correlations

<u>Correlation</u>	<u>Number of Items</u>
0.00 - 0.09	0
0.10 - 0.19	0
0.20 - 0.29	0
0.30 - 0.39	1
0.40 - 0.49	4
0.50 - 0.59	5
0.60 - 0.69	13
0.70 - 0.79	6
0.80 - 0.89	5
0.90 - 0.99	1

TABLE A9.3

Test d'arithmétique et d'algèbre de base
 Distribution of Students for Whom the Given Item
 Was the First 'Not Reached'

<u>Item</u>	<u>% of Students Not Reaching</u>	<u>Cumulative %</u>
20	1	1
21	0	1
22	0	1
23	0	1
24	2	3
25	0	3
26	1	4
27	1	5
28	1	6
29	0	6
30	0	6
31	0	6
32	0	6
33	3	9
34	1	10
35	16 ^a	26

^aFor the last item, the "not reached" figure includes also those students who reached but did not attempt the item.

TABLE A9.4

Test d'arithmétique et d'algèbre de base
Classification of Test Items by Teachers

Item	Teachers of	<u>Classification of Test Items by Teachers</u>							<u>C (College)</u> <u>D (Secondary)</u>	
		<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>		
1	Applications	11	4	0	0	0	0	0	0	0
	Fondement	12	0	0	0	0	0	0	0	0
	College	12	7	0	-	-	-	-	0	0
2	Applications	14	1	0	0	0	0	0	0	0
	Fondement	12	0	0	0	0	0	0	0	0
	College	16	3	0	-	-	-	-	0	0
3	Applications	10	5	0	0	0	0	0	0	0
	Fondement	11	1	0	0	0	0	0	0	0
	College	14	5	0	-	-	-	-	0	0
4	Applications	10	5	0	0	0	0	0	0	0
	Fondement	9	3	0	0	0	0	0	0	0
	College	12	7	0	-	-	-	-	0	0
5	Applications	6	8	1	0	0	0	0	0	0
	Fondement	8	4	0	0	0	0	0	0	0
	College	10	9	0	-	-	-	-	0	0
6	Applications	10	5	0	0	0	0	0	0	0
	Fondement	10	2	0	0	0	0	0	0	0
	College	16	2	0	-	-	-	-	1	1

TABLE A9.4 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C (College)</u> <u>D (Secondary)</u>
7	Applications Fondement College	10	4	1	0	0	0	0	0
		9	2	0	1	0	0	0	0
		14	5	0	-	-	-	-	0
8	Applications Fondement College	6	8	1	0	0	0	0	0
		5	7	0	0	0	0	0	0
		6	13	0	-	-	-	-	0
9	Applications Fondement College	12	3	0	0	0	0	0	0
		11	1	0	0	0	0	0	0
		8	7	3	-	-	-	-	1
10	Applications Fondement College	4	6	5	0	0	0	0	0
		2	6	3	0	0	1	0	0
		0	15	4	-	-	-	-	0
11	Applications Fondement College	3	8	2	0	0	0	2	0
		3	4	4	0	0	1	0	0
		0	14	5	-	-	-	-	0
12	Applications Fondement College	2	3	7	0	0	0	2	1
		1	5	5	0	0	1	0	0
		0	13	6	-	-	-	-	0
13	Applications Fondement College	7	7	0	1	0	0	0	0
		6	6	0	0	0	0	0	0
		4	11	4	-	-	-	-	0
14	Applications Fondement College	5	7	2	0	0	0	0	1
		5	4	2	0	0	0	1	0
		4	11	4	-	-	-	-	0

TABLE A9.4 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C (College)</u> <u>D(Secondary)</u>
15	Applications Fondement College	6 9 8	8 3 8	0 0 3	1 0 -	0 0 -	0 0 -	0 0 -	0 0 0
16	Applications Fondement College	6 7 3	7 5 13	1 0 3	1 0 -	0 0 -	0 0 -	0 0 -	0 0 0
17	Applications Fondement College	4 7 2	9 5 14	1 0 3	1 0 -	0 0 -	0 0 -	0 0 -	0 0 0
18	Applications Fondement College	4 7 2	9 5 14	1 0 3	1 0 -	0 0 -	0 0 -	0 0 -	0 0 0
19	Applications Fondement College	4 5 1	8 7 15	2 0 3	1 0 -	0 0 -	0 0 -	0 0 -	0 0 0
20	Applications Fondement College	3 4 2	9 6 12	2 0 3	0 0 -	0 0 -	1 1 -	0 1 -	0 0 1
21	Applications Fondement College	5 4 1	8 8 14	1 0 3	0 0 -	0 0 -	1 0 -	0 0 -	0 0 0
22	Applications Fondement College	3 2 1	10 9 13	1 0 3	0 0 -	0 0 -	1 1 -	0 0 -	0 0 1

TABLE A9.4 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C (College)</u> <u>D (Secondary)</u>
23	Applications Fondement College	8 9 1	5 3 15	0 0 3	0 0 -	0 0 -	1 0 -	0 0 -	1 0 0
24	Applications Fondement College	5 5 1	7 5 11	1 2 6	0 0 -	0 0 -	0 0 -	0 0 -	1 0 1
25	Applications Fondement College	2 1 0	4 4 8	7 7 8	0 0 -	0 0 -	0 0 -	0 0 -	2 0 3
26	Applications Fondement College	1 0 0	0 1 9	11 10 8	0 1 -	0 0 -	0 0 -	0 0 -	3 0 2
27	Applications Fondement College	9 5 2	3 6 9	0 0 7	0 0 -	1 0 -	0 0 -	0 0 -	1 1 1
28	Applications Fondement College	7 6 2	4 3 8	0 2 8	0 0 -	1 0 0	0 1 0	0 0 1	2 0 1
29	Applications Fondement College	8 6 2	3 4 8	0 2 8	1 0 -	0 0 -	0 0 -	0 0 -	2 0 1
30	Applications Fondement College	8 6 2	3 2 8	0 3 8	1 0 -	0 0 -	0 1 -	0 0 -	2 0 1

TABLE A9.4 (continued)

<u>Item</u>	<u>Teachers of</u>	<u>A1</u>	<u>A2</u>	<u>B</u>	<u>C4</u>	<u>C3</u>	<u>C2</u>	<u>C1</u>	<u>C (College) D(Secondary)</u>
31	Applications Fondement College	5 7 1	5 4 15	2 1 3	0 0 -	0 0 -	1 0 -	1 0 -	1 0 0
32	Applications Fondement College	4 7 4	6 3 10	3 2 3	0 0 -	0 0 -	1 0 -	0 0 -	1 0 2
33	Applications Fondement College	5 7 3	5 2 12	3 3 2	0 0 -	1 0 -	0 0 -	0 0 -	1 0 2
34	Applications Fondement College	3 3 1	6 4 13	2 5 5	1 0 -	0 0 -	0 0 -	1 0 -	2 0 0
35	Applications Fondement College	4 3 2	5 3 13	2 4 4	0 0 -	0 1 -	0 1 -	0 0 -	4 0 0

TABLE A9.5

Test d'arithmétique et d'algèbre de base
Description of Test Items

First Occurrence
in Guideline for

Assessment By

Item No.	Content	General		Applications		Fondement		CAAT		Difficulty
		Students	Advanced Students	Teachers	Teachers	Teachers	Teachers	Teachers	Teachers	
1	Basic arithmetic	7	7	A2	A1	A1	A2	A2	A1	0.57
2	Basic arithmetic	7	7	A1	A1	A1	A1	A1	A1	0.21
3	Basic arithmetic	8	8	A2	A1	A1	A2	A2	A2	0.49
4	Basic arithmetic	8	8	A2	A1	A1	A2	A2	A2	0.58
5	Basic arithmetic	9	9	A2	A2	A2	A2	A2	A2	0.58
6	Basic arithmetic	7	7	A2	A1	A1	A1	A1	A1	0.65
7	Basic arithmetic	7	7	A2	A1	A1	A2	A2	A2	0.58
8	Basic arithmetic	7	7	A2	A2	A2	A2	A2	A2	0.28
9	Basic arithmetic	7	7	A1	A1	A1	A2	A2	A2	0.29
10	Exponents	10	9	B	B	B	A2	A2	A2	0.23
11	Exponents	10	9	B	B	B	B	B	B	0.43
12	Exponents	11	12	B	B	B	B	B	B	0.35
13	Basic algebra	9	9	A2	A2	A2	A2	A2	A2	0.56
14	Basic algebra	9	9	A2	A2	A2	A2	A2	A2	0.67
15	Basic algebra	9	9	A2	A2	A1	A2	A2	A2	0.45
16	Basic algebra	9	9	A2	A2	A2	A2	A2	A2	0.74
17	Basic algebra	11	11	A2	A2	A2	A2	A2	A2	0.46
18	Basic algebra	10	9	A2	A2	A2	A2	A2	A2	0.51
19	Basic algebra	10	9	A2	A2	A2	A2	A2	A2	0.35
20	Basic algebra	9	9	A2	A2	A2	A2	A2	A2	0.55

TABLE A9.5 (continued)

Item No.	Content	First Occurrence in Guideline for		Assessment By				
		General Students	Advanced Students	Applications Teachers	Fondement Teachers	CAAT Teachers	Difficulty	
21	Basic algebra	9	9	A2	A2	A2	0.53	
22	Basic algebra	9	9	A2	A2	A2	0.52	
23	Basic algebra	9	9	A2	A1	A2	0.43	
24	Basic algebra	10	10	A2	A2	B	0.37	
25	Quadratic equations	12	11	B	B	B	0.30	
26	Quadratic equations	12	12	B	B	B	0.19	
27	Analytic geometry	11	10	A2	A2	B	0.27	
28	Analytic geometry	11	10	A2	A2	B	0.44	
29	Analytic geometry	11	10	A2	A2	B	0.25	
30	Analytic geometry	11	10	A2	B	B	0.31	
31	Basic algebra	9	9	B	A2	A2	0.25	
32	Basic algebra	11	11	B	A2	B	0.48	
33	Basic algebra	11	11	B	A2	A2	0.20	
34	Basic algebra	12	9	C4	B	B	0.19	
35	Basic algebra	12	11	D	B	A2	0.46	

APPENDICE B

ANALYSE DETAILLEE DES TRAVAUX ECRITS D'UN ECHANTILLON D'ELEVES FRANCO-ONTARIENS DE 12EME ET 13EME ANNEES¹

(Ce rapport a été préparé par Raymond Mougeon, Monique Bélanger et Michael Canale, Section franco-ontarienne, Institut des Etudes Pédagogiques de l'Ontario.)

1. INTRODUCTION

Dans le cadre de l'étude consacrée à la compétence en français écrit des jeunes franco-ontariens de 12ème et 13ème années, on a décidé d'analyser un sous-échantillon de 50 rédactions (50 élèves) sélectionnées à partir d'un échantillon de base (400 rédactions) afin de faire une étude détaillée de la qualité du français écrit par les élèves. La présente étude a les buts principaux suivants:

(1) pour l'échantillon de 50 rédactions:

- (a) établir pour chaque élève un indice d'erreurs basé sur toutes les erreurs commises par l'élève.
- (b) mettre en rapport ces indices d'erreurs avec les intentions professionnelles et éducatives des élèves.

(c) mettre en rapport nos indices d'erreurs avec les notes de 1 à 10 attribuées aux mêmes rédactions par un groupe d'évaluateurs.

(d) calculer des pourcentages d'erreurs ayant trait à six parties du discours particulièrement fréquentes.

(2) pour un sous-groupe de 16 rédactions sélectionné à partir de l'échantillon de 50:

(a) donner des indications détaillées sur la fréquence et l'importance des différents types d'erreurs trouvés dans les travaux écrits.

(b) proposer des explications quant à leur nature et origine.

D'une façon générale nous espérons que notre étude intéressera ceux qui, à la fois aux paliers secondaire et universitaire, sont concernés par la compétence en langue française des élèves franco-ontariens de la fin du secondaire, et qu'elle leur sera utile pour établir une liste de priorités dans le domaine de l'enseignement du français. Nous pensons également que cette étude fournira des renseignements utiles à ceux qui s'intéressent à la question du maintien du français dans les provinces du Canada où les francophones sont minoritaires. Finalement, étant donné la taille de notre échantillon, il n'est pas inutile de souligner le caractère exploratoire de notre étude. On espère que ces résultats démontreront l'utilité d'une étude approfondie basée sur l'échantillon de départ. Une telle étude permettrait sans doute de prendre connaissance de façon plus précise des difficultés linguistiques de l'ensemble des élèves franco-ontariens, ainsi que de certains des facteurs sociolinguistiques qui peuvent influencer leur maîtrise du français écrit.

2. PRESENTATION DE L'ECHANTILLON

Comme nous l'avons indiqué plus haut, le sous-échantillon de 50 rédactions a été prélevé à partir d'un échantillon de départ (400 rédactions/400 élèves). L'échantillon de départ était composé d'élèves de 12ème et 13ème années ayant reçu leur éducation en français. Ces élèves proviennent de 14 écoles franco-ontariennes (écoles unilingues et bilingues) représentant trois régions principales de concentration francophone en Ontario: le Nord, le Moyen-Nord et l'Est. Le lecteur trouvera plus de détail sur l'échantillonnage de base dans le rapport final du projet: La transition entre les niveaux secondaire et post-secondaire (projet II).

La sélection des 50 rédactions s'est faite au hasard, à trois restrictions près: il a été tenu compte (a) du sexe des élèves (25 garçons/25 filles); (b) de l'année d'étude (24 élèves de 12ème année/26 de 13ème) et (c) des sujets de rédaction choisis par les élèves (cf. Tableau B.6 pour la répartition des 50 rédactions en fonction du sujet et la liste des sujets de rédaction). Les 50 auteurs proviennent des 14 écoles franco-ontariennes mentionnées plus haut. Tous les élèves ont une connaissance active de l'anglais, mais, à des degrés différents.

Les élèves de 12ème année se répartissent dans les catégories suivantes quant à leurs projets d'avenir éducatifs et professionnels:

- (a) 13 élèves ont l'intention de continuer leurs études secondaires;
- (b) deux élèves vont aller directement dans un établissement post-secondaire;
- (c) six veulent chercher un emploi après la 12ème année et

(d) trois n'ont pas d'idées précises sur leur avenir.

Pour ce qui est des élèves de 13^{ème} année, on a obtenu la répartition suivante:

(a) 22 élèves se dirigent vers des études post-secondaires (université, collège communautaire etc.);

(b) deux élèves vont redoubler leur 13^{ème} année.

(c) un élève va chercher du travail:

(d) un élève n'a pas d'idée précise sur son avenir.

Comme on pouvait s'y attendre, la majorité des élèves de 13^{ème} année est constituée d'élèves qui vont poursuivre leurs études.

Le sous-groupe de 16 rédactions a été sélectionné au hasard sans aucune restriction. On trouvera sa répartition en fonction des principales variables mentionnées plus haut, au Tableau B.7.

3. METHODOLOGIE

La méthode utilisée dans la présente étude pour évaluer la compétence en français écrit des élèves, est connue sous le nom d'analyse d'erreurs. Il s'agit d'une méthode qui est maintenant couramment utilisée dans le domaine de l'acquisition linguistique (cf. entre autres (Corder 1976; Brown 1973; Richards 1974)). Nous l'avons nous-mêmes utilisée à plusieurs reprises (Mougeon et Hébrard 1975; Mougeon et Carroll 1976 (a) et (b); Mougeon, Bélanger, Canale et Ituen 1976) dans nos études de la compétence linguistique des jeunes Franco-ontariens de Rayside, Sudbury et Welland. Du point de vue pédagogique, l'analyse d'erreurs a le mérite principal de permettre la réalisation d'études diagnostiques détaillées de la compétence linguistique des

"apprenants" d'une langue donnée, études qui peuvent avoir des retombées dans le domaine de l'enseignement des langues (Burt et Kiparsky 1972; Mougeon 1975). Il convient néanmoins de signaler un problème majeur posé par la méthode de l'analyse d'erreurs, problème qui a trait à la notion d'erreur. Plus précisément la notion d'erreur est étroitement liée à la notion de norme, l'erreur étant ce qui ne se conforme pas à la norme. Il en résulte que le choix d'une norme linguistique donnée, à des fins d'analyse, conditionne en partie l'évaluation de la compétence linguistique d'un individu. D'une façon générale nous pensons qu'il est possible d'apporter une solution partielle à ce problème si, d'une part on choisit une norme linguistique qui correspond au type de comportement linguistique (formel, informel, écrit, parlé, etc) que l'on attend d'un individu dans une ou des situation(s) de communication donnée(s) et si d'autre part, on définit clairement les caractéristiques de la norme adoptée à des fins d'analyse. Dans le cadre de la présente étude nous avons décidé d'adopter le français canadien formel écrit comme norme de référence. Ce choix est motivé par l'orientation de l'étude générale dans laquelle s'inscrit notre étude. En effet, la première a notamment pour but de voir si les élèves ontariens de la fin du secondaire possèdent les capacités requises pour poursuivre des études au niveau post-secondaire. Plus précisément, nous sommes partis du principe que dans le cadre des études post-secondaires on attendra généralement des élèves qu'ils maîtrisent une variété de langue conforme aux règles d'usage du français canadien soutenu. Ceci dit, nous avons choisi de façon un peu arbitraire de nous baser sur l'usage qui est décrit dans le dictionnaire du français canadien (Bélisle 1971) dans la mesure où cet ouvrage propose un modèle explicite de français canadien "correct". Il n'en reste pas moins que le dictionnaire Bélisle ne recouvre pas tous les aspects de l'usage de la langue française, notamment pour ce qui est de la grammaire. Nous avons donc dû parfois avoir recours à nos propres intuitions au sujet de la norme du français canadien formel.² Moyennant cette norme de référence, nous avons recueilli en examinant tous les travaux écrits, tous les éléments qui ne se conformaient pas à la norme (erreurs) dans les domaines principaux de l'orthographe, du

vocabulaire et de la grammaire. La collecte des erreurs a été effectuée par deux linguistes (Monique Bélanger et Raymond Mougeon) qui ont corrigé ensemble chacune des rédactions. Ainsi a-t-on obtenu un nombre d'erreurs plus élevé qu'il n'est possible d'en obtenir quand une personne seule s'adonne à la tâche ingrate de la collecte des erreurs. D'autre part, nous avons réduit un tant soit peu le risque d'évaluation subjective, dans les cas où nous avons eu affaire à des éléments linguistiques non répertoriés dans le dictionnaire Bélisle.

4. RESULTATS

Commençons par les indices d'erreurs trouvés pour chacun des 50 élèves. Avant de présenter les résultats sous forme de tableaux, il convient de donner quelques explications au sujet de nos indices d'erreurs. Ces derniers ont été calculés en divisant le nombre total d'erreurs commises par un élève donné par le nombre total de mots écrits par l'élève. Cette méthode a déjà été utilisée par Scott et Tucker (1974), et Mougeon et Hébrard (1975). Un tel indice ne peut nous donner qu'une indication générale sur la maîtrise du français écrit par les élèves, ceci pour plusieurs raisons: (a) il peut y avoir plus d'une faute par mot (cf. plus bas); (b) tous les mots ne présentent pas le même niveau de difficulté; (c) tous les étudiants ne commettent pas les mêmes types d'erreurs dans les mêmes proportions. Moyennant ces réserves, nos indices sont d'un intérêt appréciable, dans la mesure où ils représentent une mesure relativement objective de la compétence linguistique des élèves, mesure qui repose sur la totalité des erreurs commises par ces derniers.

Comme l'indique le Tableau B.1, les deux groupes principaux d'élèves de 12^{ème} année (ceux qui vont continuer leurs études et ceux qui vont chercher du travail) ont des capacités à écrire le français nettement différentes. Plus précisément, on constate une différence de plus de 10 points entre l'indice moyen du premier groupe et celui du deuxième groupe. Il est aussi intéressant de

remarquer, que l'indice le plus élevé trouvé pour le premier groupe, correspond à peu de chose près à l'indice le plus bas trouvé pour le deuxième groupe. On notera également, que la minorité qui se dirige directement vers une institution post-secondaire, a également un indice d'erreurs bas. L'existence d'une relation entre les intentions professionnelles et éducatives des 50 élèves et leurs capacités à écrire le français (telles que mesurées par notre indice), est un fait intéressant que nous croyons bon de souligner. En effet, on peut supposer qu'un nombre appréciable des élèves qui déclarent vouloir continuer leurs études secondaires, se dirigeront sans doute après la 13^{ème} année vers un établissement post-secondaire. Comme nous l'avons indiqué plus haut, dans de tels établissements, on attend de l'élève une bonne connaissance de la langue écrite.

Parmi les élèves de 13^{ème} année (Tableau B.2) on trouve surtout des élèves qui ont l'intention de faire des études dans un établissement post-secondaire; les scores de ce groupe d'élèves présentent donc un intérêt tout particulier. On constatera d'abord, que l'indice moyen de ce groupe est proche de celui du groupe des 12^{ème} année qui ont l'intention de continuer leurs études. Toutefois, on se doit aussi de remarquer que contrairement à l'attente normale, nous avons obtenu une large distribution pour les scores du sous-groupe d'élèves de 13^{ème} année qui veulent continuer leurs études (de .01 à .22). En effet on aurait pu s'attendre à trouver parmi ce sous-groupe une nette majorité d'élèves ayant des indices d'erreurs bas ou relativement bas étant donné les intentions éducatives de ces élèves. Or on constate que seulement une faible majorité des élèves de ce groupe ont des indices inférieurs à .10. Sur la base de ces résultats, on peut supposer qu'un certain nombre des élèves de 13^{ème} année qui veulent continuer leurs études et qui ont des indices d'erreurs assez élevés, éprouveront des difficultés certaines s'ils ont à étudier dans un établissement post-secondaire de langue française. A ce sujet, il serait intéressant de mesurer la compétence en anglais écrit de ces élèves afin de voir si ceux-ci auraient avantage à faire leurs études en anglais. Etant donné la taille très réduite des autres

sous-groupes d'élèves de 13^{ème} année, nous nous abstiendrons de faire des commentaires au sujet des scores de ces élèves.

Il reste une question importante à laquelle il nous est difficile de répondre: celle de savoir quel est le niveau de maîtrise du français écrit exigé de façon générale par les établissements post-secondaires de langue française. Faute de pouvoir répondre à cette question, nous ne pouvons pas indiquer dans quelle proportion les élèves de notre échantillon de 50, possèdent la compétence en français écrit nécessaire aux études post-secondaires.

Passons maintenant à la comparaison de nos indices avec les notes globales données par les évaluateurs. A ce sujet, il faut mentionner que chaque rédaction a été évaluée par neuf juges différents. Ceux-ci ont eu pour consigne de noter chaque rédaction de un à dix en se basant sur leur impression générale. Ils devaient également passer en moyenne deux minutes par rédaction. Il nous est donc apparu intéressant de confronter ces deux méthodes d'évaluation différentes. Pour ce faire, nous avons essayé de voir s'il existe une corrélation entre la note moyenne obtenue par chaque élève et l'indice d'erreurs correspondant. En appliquant le test "Pearson Product Moment" nous avons trouvé une corrélation relativement élevée (-74). Le fait que la corrélation soit négative indique que plus les notes des juges sont élevées, moins les indices d'erreurs le sont. Nous avons illustré cette corrélation de façon graphique (cf. graphique 1). Ce résultat nous autorise à conclure que la méthode de l'évaluation globale et rapide a abouti à des résultats similaires à ceux obtenus par l'analyse détaillée et quasi exhaustive des erreurs. On peut rapprocher ceci du résultat semblable trouvé par P. Evans qui a étudié les rédactions de l'échantillon d'élèves anglo-ontariens. On peut donc voir dans ces résultats une indication que la méthode de l'évaluation globale et rapide peut représenter un raccourci relativement fiable. Il n'en reste pas moins (ceci est évident) que la méthode de l'évaluation globale ne permet pas de diagnostiquer avec précision l'importance et la fréquence respective des différents types d'erreurs commises par les élèves.

Examinons maintenant les pourcentages d'erreurs calculés sur la base des 50 rédactions. Comme ces calculs prennent énormément de temps, nous nous sommes limités à six parties du discours ayant une fréquence d'occurrence élevée. Pour chacune de ces parties du discours (excepté les prépositions), nous avons analysé deux types d'erreurs: celles qui ont trait à l'usage lexical et celles qui ont trait à l'accord de nombre. Les pourcentages d'erreurs ont été calculés en divisant le nombre total d'erreurs trouvé pour une partie du discours donnée dans les 50 rédactions, par une estimation de la totalité des occurrences de cette même partie du discours. Pour arriver à une telle estimation, nous avons compté le nombre total d'occurrences d'une partie du discours donnée dans un sous-groupe de 15 rédactions sélectionnées au hasard. Ceci nous a permis d'établir la fréquence moyenne d'occurrence; avec cette fréquence nous avons estimé le nombre total d'occurrences des six parties du discours pour les 50 rédactions.

Commençons par examiner les erreurs d'usage (Tableau B.3)

Avant de commenter les résultats concernant les erreurs d'usage, nous allons définir brièvement ce que nous entendons par erreur d'usage et fournir ensuite une série d'exemples tirés des 50 rédactions. Faire une erreur d'usage consiste à ne pas utiliser ce qu'on appelle communément le mot juste. Les erreurs d'usage sont d'une gravité variable, dans la mesure où elles peuvent nuire plus ou moins à la communication linguistique. Dans notre étude, elles recouvrent deux catégories principales: l'omission d'un élément linguistique ou la substitution d'un élément linguistique à un autre. Elles incluent aussi (dans des proportions moindres) l'utilisation d'un élément linguistique dans un contexte où il n'est pas requis.

Présentons maintenant quelques exemples illustrant les différents types d'erreurs qui figurent au Tableau B.3.

- (a) Prépositions: (1) S'il y a une partie de balle___quelques blocs d'ici (omission) (2) Le Canada a beaucoup à offrir pour les jeunes (substitution) (3) L'étudiant n'a pas de besoin de ces cours (utilisation non requise)
- (b) Verbes: (1) Sans ce stade aucun succès significatif pourra être aperçu (substitution) (2) Il y a des aides du gouvernement pour fournir aux besoins de tous (substitution) (3) En réflétant sur le passé, l'homme inventa une communication très utile (substitution)
- (c) Substantifs: (1) Pour atteindre le pas de la compétition (substitution) (2) La timidité, la non-confiance...décourage souvent les compétiteurs (substitution) (3) L'homme inventa une communication très utile (substitution)
- (d) Article: (1) Il construit___machines à détruire le monde (omission) (2) ... avec___I.Q. de 130 (omission) (3) Le monde de communication est un spectre qui permet aux gens de faire face à une dure réalité qui nous entoure (substitution)
- (e) Adjectif: (1) Sans ce stade aucun succès significatif pourra être aperçu (substitution) (2) La première journée de ta naissance (utilisation non requise)
- (f) Pronom: (1) On a tendance à ___humilier (omission) (2) Par soi, il désire en connaître plus encore (substitution) (3) ___sont jamais vrais (omission)

Le tableau B.3 indique que les six parties du discours créent plus ou moins de difficultés aux étudiants. On remarquera que ces derniers semblent avoir des problèmes relativement sérieux avec les prépositions et avec les verbes. Pour ce qui est des prépositions, on peut signaler que plusieurs études (Oller et Inal 1971; Scott et Tucker 1974; Mougeon et Hébrard 1975;

Mougeon et Carroll 1976) ont montré que les "apprenants" d'une langue première ou seconde éprouvent généralement des difficultés à maîtriser les systèmes de prépositions. A notre avis, ceci est dû en partie au fait que, bien que les prépositions forment un ensemble d'éléments restreint, elles prennent, en se combinant aux verbes et aux substantifs, une multitude de sens différents qui ne sont pas toujours généralisables. Ceci dit, il est également possible de considérer les prépositions comme des éléments qui ne sont pas absolument essentiels à la communication (Brown 1973) dans la mesure où les erreurs de préposition n'entravent généralement pas la compréhension du message linguistique. Ceci explique peut-être en partie pourquoi, parmi les erreurs de prépositions, on trouve une proportion non négligeable d'omissions. Les verbes, par contre, figurent parmi les éléments essentiels du message linguistique. Le pourcentage relativement élevé d'erreurs trouvé pour ceux-ci, nous porte à recommander que l'on mette l'accent sur leur usage. Le fait que les étudiants éprouvent moins de problèmes avec les substantifs qu'avec les verbes, constitue un résultat inattendu de notre étude. Nous aimerions voir s'il est confirmé par une étude qui serait basée sur un échantillon plus grand. Le fait que les élèves tendent à avoir une meilleure maîtrise des adjectifs que des verbes est à souligner aussi, dans la mesure où il va plus ou moins à l'encontre d'une "ancienne" théorie récente (Jacobs et Rosenbaum 1968) qui postulait que les adjectifs sont sémantiquement analogues aux verbes. Pour ce qui est des pourcentages d'erreurs relativement bas trouvés pour les articles et les pronoms, on peut l'expliquer par le fait que les systèmes des pronoms et des articles nécessitent l'acquisition d'un nombre restreint de règles qui (contrairement aux prépositions) sont hautement généralisables, bien que dans certains cas l'utilisation des articles (opposition défini/indéfini) puisse être assez subtile.

Passons maintenant aux erreurs ayant trait à l'accord de nombre. Le tableau B.4 ne concerne que cinq parties du discours. Les prépositions, qui ne sont pas sujettes à l'accord de nombre, ont été exclues.

Par erreur de nombre nous entendons l'emploi de la marque du pluriel là où un singulier est requis et vice versa. Donnons-en quelques exemples tirés des 50 rédactions.

- (a) ADJECTIF: (i) Ils ont des contacts physique__violent__. (ii) Ils mangent à leurs faim. (III) Comparativement à d'autre__ __pays sous-développé__
- (b) VERBE: (i) Elles se situe__loin dans le temps. (ii) Ceux qui aime__la nature. (iii) Ces machines fait__par l'homme.
- (c) SUBSTANTIF: (i) Son amis__ __lui apprend du nouveaux. (ii) L'enfant passe à travers beaucoup d'échec__. (iii) De jours en jours.
- (d) PRONOM: (i) Seulement pour toi et non pour d'autre__. (ii) Tous ce qui l'entoure. (iii) Le genre de vie qui leurs plait.
- (e) ARTICLE: (i) Les temps de__dépressions et de__guerres. (ii) Il est nécessaire de suivre de__ __cours spécifiques.

On notera d'abord que l'article n'est pratiquement pas affecté par les erreurs de nombre. Ceci est sans doute dû en partie au fait qu'en français la marque pluriel/singulier est à la fois essentielle et évidente au niveau de l'article. Plus précisément, si l'on considère les deux phrases: les petits garçons restent tranquilles et le petit garçon reste tranquille, on s'aperçoit qu'en langue parlée la différence pluriel/singulier n'est indiquée que par l'alternance entre les articles les et le, d'où sans doute le fait que nous ayons trouvé très peu d'erreurs de nombre au niveau de l'article. Si l'article est très peu affecté par les erreurs de nombre ceci est aussi vrai (à peu de choses près) pour les substantifs et les pronoms personnels. En ce qui concerne les substantifs, ceci s'explique peut-être en partie

par la relation de solidarité qui existe entre le substantif et l'article. Pour ce qui est des pronoms personnels, on notera que dans la plupart des cas la différence entre le singulier et le pluriel est évidente, sauf pour les pronoms de la 3ème personne il/ils, elle/elles, leur/leurs, autre/autres qui ne sont pas différenciés en langue parlée. C'est du reste ces derniers pronoms qui ont surtout fait l'objet des erreurs de nombre. Pour ce qui est des verbes, on notera que, en dépit du fait que la différence de nombre est généralement indiquée de façon évidente en langue parlée par l'alternance des terminaisons (e/ont, -s/ez etc.), il y a un risque de confusion avec les terminaisons de la 3ème personne e/ent et ait/aient qui sont prononcées de façon indentique. C'est du reste en grande partie au niveau de ces dernières terminaisons que nous avons trouvé des erreurs de nombre. On notera également que les erreurs de nombre ont affecté les participes passés (cf. exemple b(iii)). Finalement, en ce qui concerne les adjectifs, on notera qu'à l'inverse des substantifs, ils sont souvent éloignés de l'article, exception faite de la minorité des adjectifs qui sont utilisés avant le nom. Ceci explique peut-être pourquoi nous avons trouvé un taux relativement plus élevé d'erreurs de nombre les affectant.

Passons maintenant à l'analyse détaillée des différents types d'erreurs. Comme nous l'avons indiqué plus haut cette étude est basée sur un sous-groupe de 16 rédactions (cf Tableau B.7) et non sur les 50 rédactions, car nous n'avons pas disposé d'assez de temps.

Le tableau B.5 fait apparaître dix types d'erreurs différents. Ceux-ci ont été ordonnés en fonction de leur importance respective, à l'intérieur de l'ensemble de toutes les erreurs trouvés dans les 16 rédactions. Ces pourcentages ne nous donnent qu'une indication assez grossière de l'ampleur des difficultés que crée respectivement chacun des types d'erreurs aux élèves. Ceci est dû au fait que les types d'erreurs portent sur des éléments linguistiques de fréquence variable et que la variation de cette fréquence se répercute plus ou moins au niveau du nombre absolu d'erreurs d'un type donné. Idéalement, nous

aurions dû calculer pour chacun de ces dix types d'erreurs, des pourcentages d'erreurs du type de ceux que nous avons présentés aux tableaux B.3 et B.4, mais étant donné l'ampleur de la tâche, nous avons dû y renoncer. Moyennant ces réserves, on peut noter que la majorité des erreurs est couverte par quatre types d'erreurs. Deux de ceux-ci (usage lexical, accord de nombre) ont été examinés rapidement dans la section précédente. On voit que la catégorie des erreurs d'orthographe arrive en deuxième. Elle devrait d'après nous, être une des principales préoccupations des responsables de l'enseignement du français, d'autant plus que nous n'avons pas compté certaines erreurs d'orthographe mineures (cf. plus bas). Le reste des erreurs consiste en un groupe de types d'erreurs de fréquence assez peu élevée, mis à part le genre, qui semble créer des difficultés aux étudiants.

Ceci dit, examinons plus en détail ces différents types d'erreurs, en particulier ceux qui n'ont pas été examinés dans la section précédente. Commençons par un des types les plus importants, à savoir les erreurs d'orthographe. Indiquons à ce sujet que nous n'avons pas compté les erreurs d'accentuation, sauf quand l'accent correspond à une différence de fonction (exemple a/à, ou/où). Ont été également exclues, les erreurs de ponctuation (points sur les "i", virgules, etc.) et celles portant sur l'utilisation des lettres capitales. Nous avons différencié deux types d'erreurs d'orthographe, celles attribuables à l'homonymie et celles qui ne le sont pas. Nous avons identifié 77 erreurs (44%) qui découlent de l'homophonie et 97 (57%) qui n'en découlent pas. La première catégorie recouvre 17 différents types de conflits homonymiques (exemples: sont/son, ce/se, etc.). Ceux-ci sont listés au tableau B.8. Nous les avons ordonnés en fonction de leur fréquence.

L'erreur consiste à utiliser une forme à la place d'une autre. On notera en particulier parmi les erreurs les plus fréquentes celles qui ont trait aux paires --é/--er (de loin la plus fréquente), à/a, qui/qu'il, ce/se et aux trois doublets s'est/c'est; ces/ses; sait/sais. Voici quelques exemples de certaines de ces erreurs: Je suis d'accord que ses trois causes

sont responsables; et on a tendance a humilié ses ennemis; et se seras impossible de se separée de la violence; nous voulons être nous-même, et fair ce qu'il nous plait. A notre avis, ces erreurs découlent en grande partie d'un mauvais apprentissage des différentes catégories grammaticales du français, qui fait que l'élève ne sait plus différencier les participes des infinitifs, les démonstratifs des possessifs, etc. Or ce discernement est une des clefs essentielles de l'orthographe française, dans la mesure où cette dernière est malheureusement beaucoup plus grammaticale et étymologique que phonétique. A ce sujet, nous renvoyons le lecteur à Martinet (1969), s'il veut approfondir cette question.

La deuxième catégorie d'erreurs, comporte en majeure partie des erreurs qui ont trait au doublement des consonnes et, dans une moindre mesure, à la simplification des groupes de lettres qui ne sont pas prononcés comme ils sont écrits. En voici quelques exemples: une communication se fesait en quelques secondes; un problème sérieux celui de la polution; détruire des milliers de bactéries; tout en prennant un bref aperçu de la situation; la libération de la famme.

En ce qui concerne les erreurs d'usage lexical (le plus fréquent des dix types d'erreurs), signalons qu'en plus des substantifs, verbes, adjectifs et prépositions (examinés plus haut), elles ont aussi porté sur les adverbes, les pronoms et les conjonctions. On trouvera au tableau B.9 des renseignements sur l'importance respective des différents types d'erreurs d'usage. A ce sujet, on peut remarquer que les erreurs d'usage ont été divisées en deux catégories; celles qui sont plus ou moins attribuables à l'influence du lexique anglais (erreurs d'interférence) et celles qui ne le sont pas. Voici quelques exemples d'erreurs d'interférence: on nous dit par après que le blamme est sur la télévision; la roue nous a élevé d'un stage primitif a un stage de progrès; sur le côté humanitaire, on peut vivre plus longtemps...; la transportation est un autre domaine qui ... etc.

Il n'est peut-être pas inintéressant de noter que nous n'avons trouvé des erreurs d'interférence qu'avec les substantifs, les verbes, les adjectifs et les prépositions, et ce dans des proportions relativement similaires pour chacune de ces parties du discours, exception faite des prépositions qui semblent avoir été un peu plus l'objet d'erreurs d'interférence. Un fait similaire a déjà été noté par Mougeon et Hébrard (1975) et Mougeon et Carroll (1976(c)) relativement à l'acquisition des prépositions anglaises par des jeunes franco-ontariens de Sudbury et Welland. A notre avis, la complexité des systèmes des prépositions françaises et anglaises n'est sans doute pas étrangère au fait, qu'à la fois en anglais et en français, les bilingues commettent de nombreuses erreurs de prépositions dues à l'interférence. Plus précisément, ayant à faire face aux tâches complexes d'acquérir et d'utiliser deux systèmes de prépositions, les élèves bilingues plus ou moins consciemment simplifient ces tâches, en transférant certaines des règles d'usage d'un système dans l'autre et vice versa. Ce faisant, les élèves produisent des structures non conformes aux normes du français et de l'anglais canadiens telles qu'utilisées par les unilingues (cf. Mougeon et Carroll, 1976b; Mougeon, Bélanger, Canale et Ituen, 1976).

Pour ce qui est des erreurs de nombre, les ayant déjà examinées plus haut, signalons seulement que si l'erreur de nombre consiste à utiliser le singulier pour le pluriel et vice versa, la substitution du singulier au pluriel s'est avérée plus fréquente (60%) que la substitution inverse (40%). Une telle tendance a déjà été constatée dans plusieurs études consacrées à l'acquisition de l'anglais (première langue: Brown, 1973, langue seconde: Mougeon et Hébrard, 1975) et de l'espagnol (Chun et Politzer, 1975). Il faut y voir sans doute l'existence d'une tendance à surgénéraliser la forme qui est morphologiquement plus simple (cf. plus bas les erreurs de genre).

Les erreurs de genre ont surtout porté sur quatre parties du discours: les articles et les substantifs, les adjectifs et les participes passés. Ces deux dernières parties du discours recouvrent plus de 70% des erreurs de genre. Cette proportion est

d'autant plus élevée que les adjectifs et les participes passés ne sont pas des items très fréquents. Il se peut donc que l'accord de genre soit perçu par les élèves comme nettement redondant au niveau de l'adjectif et du participe passé, dans la mesure où le genre apparaît en premier et de façon évidente (cf. plus haut) au niveau du groupe article-substantif. Voici quelques exemples d'erreurs de genre: comme étudiantant active dans les sports; ceci est une accomplissements très importantes; une bonne idée est souvent apprécié; ce n'est pas l'honneur individuelle que ces athlètes recherchent. Comme on peut le voir ci-dessus, l'erreur consiste à substituer le féminin au masculin et vice-versa. De même que pour le nombre, nous avons trouvé que la substitution de la forme plus simple (masculin) à la forme plus complexe (féminin) est légèrement plus fréquente que la substitution inverse. On peut rapprocher ce résultat des découvertes similaires faites par Grégoire (1947) (français langue première), Swain (1975), Tarone, Frauenfelder et Selinker (1975) (français langue seconde). La plupart des erreurs concernant la personne du verbe ont trait aux marques de la deuxième et de la troisième personnes du singulier. Celles-ci sont, soit omises, soit substituées l'une à l'autre. En voici quelques exemples: même si nous venons à abolir les films de violences, notre société serais dôle. Ce seras impossible de se séparée de la violence. En grandissant tu connaîtra la définition de ce mot; il essais de se tenir sur ses deux pieds.

Les erreurs d'article sont uniquement des erreurs d'usage. Elles se divisent en deux catégories principales, l'omission des articles définis, indéfinis et partitifs (la majorité des erreurs d'articles), la substitution d'un article défini à un article indéfini et vice-versa. Le lecteur trouvera des exemples de ces erreurs dans une section précédente.

Dans la catégorie des erreurs de syntaxe on trouve surtout des erreurs qui ont trait à l'ordre des éléments dans la phrase, Exemples: mettra t-il à fin la pauvreté? Pourquoi devrait un étudiant prendre un cours de français? Tout de même, dans ce monde de moderne technologie on peut...

On remarquera que les élèves ont, dans l'ensemble, peu de problèmes avec l'ordre des éléments dans la phrase puisqu'on a trouvé peu d'erreurs de ce genre. La même remarque est valable pour l'omission de la particule négative ne, exemple: il__veut pas les écouter; cette dernière est rarement omise par les étudiants. Ce résultat est intéressant car une étude récente (Vincent & Sankoff, 1975) a montré qu'en français canadien parlé, la particule ne est omise dans plus de 95% des cas. On peut donc remarquer que les étudiants n'ont pratiquement pas transféré dans leur écrit l'omission de ne, caractéristique du français parlé.

La catégorie des erreurs appelées "autres" recouvre différents types d'erreurs de basse fréquence. Parmi les plus importantes on peut citer l'omission du pronom personnel, exemple: __faut mettre en considération le fait que...; l'omission du relatif que, exemple: je pense__je vais faire un ingénieur; l'utilisation redondante de que, exemple: quand qu'ils se trompent ils perdent confiance. On notera que ces erreurs peuvent être expliquées en partie comme des transferts de traits caractéristiques du français canadien parlé.

Finalement ont été classées comme erreurs non analysables celles pour lesquelles il s'est avéré impossible de statuer de façon plausible sur leur nature et origine. Dans cette catégorie nous trouvons en grande majorité des phrases incomplètes et difficilement interprétables. En voici quelques exemples; En fait tout y est la participation et enfin la victoire ou la défaite; Certainne personne en voyent des films ou la télévision avec violence croit que ça peut leur arriver; Il s'étonne pour les livres permettent la distribution à une âge. Il s'agit comme on peut le constater d'erreurs assez graves dans la mesure où elles entravent la compréhension du message linguistique. Pour cette raison, nous aurions aimé en trouver un nombre infime dans les rédactions des élèves.

Avec ce dernier type d'erreurs, s'achève la section consacrée aux erreurs du sous-échantillon de 16 rédactions. Dans la section suivante, nous allons examiner les limites de notre

étude suggérant par là même des dimensions nouvelles pour une étude ultérieure basée sur un échantillon plus large.

5. CONCLUSION

Signalons tout d'abord que l'analyse des différents types d'erreurs trouvés dans le sous-échantillon de 16 rédactions fait abstraction des différences individuelles qui parfois peuvent être frappantes. Ainsi nous avons remarqué que certains élèves tendaient à faire plus d'erreurs d'un type que d'un autre. Certains élèves, par exemple, font surtout des erreurs d'orthographe et plus particulièrement des erreurs d'accord; d'autres ont un vocabulaire très approximatif; d'autres, et cela est plus grave, produisent un nombre relativement élevé de structures incomplètes et obscures. Une étude basée sur un échantillon plus vaste devrait à notre avis tenir compte de ces différences individuelles. Signalons également que dans la présente étude, nous n'avons pas abordé la question épineuse des écarts stylistiques. Ainsi certains élèves ont utilisé une quantité non négligeable de structures caractéristiques de la langue parlée informelle. Parmi celles-ci, on peut citer le redoublement du pronom, exemple: moi je; lui il; il ne le comprend pas ça, etc.; l'interrogation avec tu, exemple: il comprend tu ce que le monde veulent; l'utilisation redondante de y, exemple: quand il y allait à l'école. Nous n'avons pas compté ces structures comme des erreurs car ce sont des points relativement mineurs, toutefois, il faut admettre qu'elles indiquent qu'un certain nombre d'élèves n'ont pas une bonne maîtrise du français formel écrit (cf. Mougeon et Carroll, 1976 (a) et (b) pour des remarques similaires). Dans la mesure où on peut supposer que les établissements post-secondaires s'attendent à ce que les étudiants fassent preuve d'une telle maîtrise, on ne saurait trop recommander que l'on sensibilise les étudiants de la fin du secondaire à l'existence des niveaux de style. Ceci nous semble d'autant plus impératif que de par leur statut minoritaire, les jeunes Franco-ontariens sont assez peu exposés aux variétés de français formel. En effet, on peut

rappeler que la création d'un enseignement de langue française au niveau secondaire est chose récente. Il en est de même pour la télévision et la radio françaises (deux véhicules de variétés de français formel). A cela, on peut ajouter que dans de nombreuses localités où les Franco-ontariens sont présents, il est difficile de se procurer des livres et des journaux français.

Finalement, mentionnons que nous n'avons pas abordé la question importante de la logique dans la présentation des idées, dans la mesure où elle déborde le cadre de l'analyse linguistique. Cependant, signalons que seule une infime minorité des étudiants a fait preuve d'une telle capacité. En fait, la plupart du temps, nous avons constaté que les idées sont mal enchainées. Pourtant, étant donné la nature des thèmes des rédactions, on aurait pu s'attendre à ce que les élèves suivent une démarche démonstrative. Etant donné que dans le contexte des études post-secondaires les étudiants auront sans doute à fonctionner de façon discursive, on devrait peut-être mettre davantage l'accent sur l'entraînement au discours logique dans les classes terminales des écoles secondaires.

TABLEAU B.1

Indices d'erreurs des élèves de 12ème année en fonction de leurs intentions professionnelles et éducatives

<u>Intentions des élèves</u>	<u>Ss</u>	<u>Total des mots écrits par l'élève</u>	<u>Total d'erreurs commises par l'élève</u>	<u>Indices d'erreurs</u>
Vont	1	700	17	.02
	2	420	11	.03
continuer	3	443	18	.04
	4	360	22	.06
leurs	5	445	32	.07
	6	300	31	.10
études	7	337	32	.10
	8	562	67	.12
secondaires	9	460	60	.13
	10	256	33	.13
	11	436	57	.13
	12	641	86	.13
	13	391	55	.14
Total		<u>5751</u>	<u>521</u>	<u>.09</u>
<hr/>				
Vont directe-	1	400	24	.06
ment dans une	2	<u>740</u>	<u>58</u>	<u>.08</u>
institution				
post-secondaire				
Total		1140	82	.07
<hr/>				
Vont	1	176	22	.13
	2	344	54	.16
chercher	3	436	85	.19
	4	397	87	.22
du travail	5	611	132	.22
	6	<u>348</u>	<u>95</u>	<u>.27</u>
Total		<u>2312</u>	<u>475</u>	<u>.20</u>
<hr/>				
N'ont pas	1	457	56	.12
d'idée	2	168	22	.13
précise	3	<u>524</u>	<u>83</u>	<u>.16</u>
Total		<u>1149</u>	<u>161</u>	<u>.14</u>

TABLEAU B.2

Indices d'erreurs des élèves de 13ème année en fonction de leurs intentions professionnelles et éducatives

<u>Intentions de l'élève</u>	<u>Ss</u>	<u>Total des mots écrits par l'élève</u>	<u>Total d'erreurs commises par l'élève</u>	<u>Indices d'erreurs</u>
Vont	1	753	8	.01
	2	449	20	.04
	3	555	30	.05
dans	4	507	28	.06
	5	324	24	.07
une	6	742	54	.07
	7	355	24	.07
institution	8	615	43	.07
	9	488	41	.08
post-	10	594	50	.08
	11	337	29	.09
secondaire	12	372	42	.11
	13	417	49	.12
	14	502	62	.12
	15	289	38	.13
	16	562	73	.13
	17	386	57	.15
	18	756	128	.17
	19	444	75	.17
	20	483	97	.20
	21	191	43	.22
Total		10121	1025	.10
<hr/>				
Vont refaire	1	498	43	.09
une 13ème	2	405	69	.17
Total		903	112	.12
<hr/>				
Vont chercher du travail	1	616	15	.02
<hr/>				
N'ont pas d'idée	1	406	27	.07
précise		287	49	.17
Total		793	76	.09

TABLEAU B.3

Pourcentages d'erreurs d'usage pour
six parties du discours fréquentes

<u>Item</u>	<u>Nombre d'erreurs</u>	<u>Estimation du nombre total d'item</u>	<u>Pourcentage d'erreurs</u>
Préposition	268	2700	10%
Verbe	193	2700	7%
Substantif	152	4800	" 3%
Article	79	2700	3%
Adjectif	52	1800	3%
Pronom	46	1800	2%

TABLEAU B.4
 Pourcentages d'erreurs de nombre pour
 cinq parties du discours fréquentes

<u>Item</u>	<u>Nombre d'erreurs</u>	<u>Estimation du nombre total d'items</u>	<u>Pourcentage d'erreurs</u>
Adjectif	106	1800	6
Verbe	140	2700	5
Substantif	147	4800	3
Pronom personnel	50	1800	3
Article	33	2700	1

TABLEAU B.5

Les différents types d'erreurs
trouvées dans les 16 redactions

<u>Type</u>	<u>Nombre d'erreurs</u>	<u>Pourcentage de fréquence</u>
Usage lexical	237	32%
Orthographe	174	23%
Nombre	154	20%
Genre	42	6%
Personne du verbe	29	4%
Article	23	3%
Erreurs de syntaxe	11	1%
Omission de ne	10	1%
Autres	50	7%
Erreurs non-analysables	21	3%
Total	751	100%

TABLEAU B.6

Repartition des 50 redactions en fonction du thème

<u>Thème</u>	<u>Nombre d'élèves</u>
Thème 1	10
Thème 2	5
Thème 3	5
Thème 4	10
Thème 5	5
Thème 6	5
Thème 7	5
Thème 8	5

Thème 1: Les scènes de violence dans les livres, les films ou à la télévision doivent-elles être censurées?

Thème 2: Quelles sont les valeurs des Jeux Olympiques, non seulement pour les concurrents, mais plus particulièrement pour les pays qu'ils représentent et pour le monde entier?

Thème 3: Le progrès de la science est la garant du progrès de la civilisation.

Thème 4: Les écoles secondaires ne devraient pas demander que les étudiants suivent tel ou tel cours ou aient tel ou tel pré-requis; les étudiants devraient être libres de choisir les cours qui correspondent le mieux à leurs intérêts.

Thème 5: La compétition développe la force de caractère.

Thème 6: Les années 70 sont les meilleures pour être jeune au Canada.

Thème 7: Une des caractéristiques de notre temps est notre réticence à participer; nous devenons de plus en plus une nation de spectateurs.

Thème 8: L'art imite-t-il la vie, ou la vie imite-t-il l'art?

TABLEAU B.7

Repartition du sous-groupe de 16 élèves en fonction de leurs intentions et de l'année d'étude

<u>Vont continuer leurs études secondaires</u>		<u>Vont dans un établissement post-secondaire</u>		<u>Vont chercher du travail</u>	<u>N'ont pas d'idée</u>	
<u>12ème</u>	<u>13ème</u>	<u>12ème</u>	<u>13ème</u>	<u>12ème</u>	<u>12ème</u>	<u>13ème</u>
1	2	2	7	2	1	1

TABLEAU B.8

Détail des différentes erreurs d'orthographe

Erreurs attribuables aux conflits homonymiques

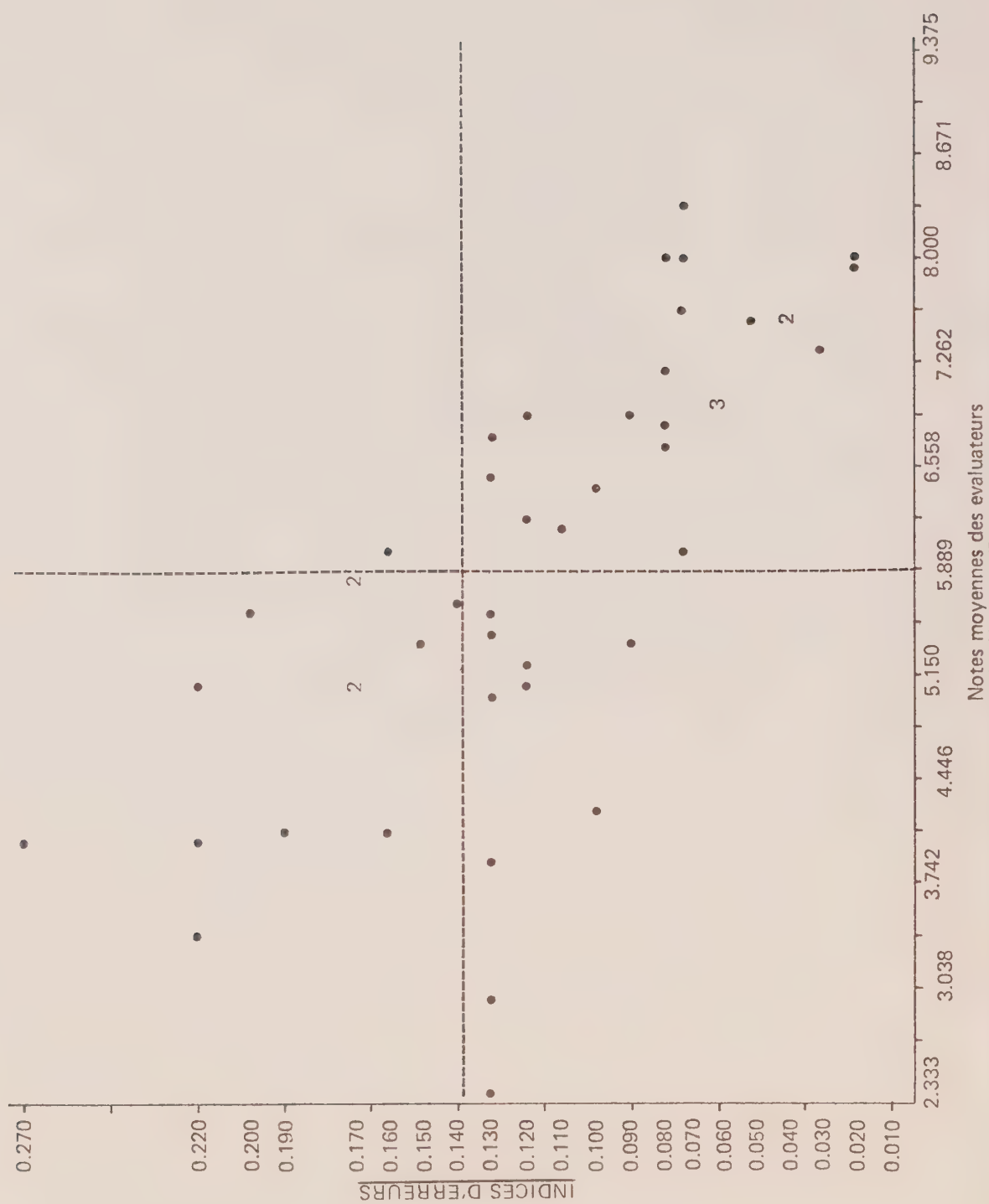
<u>Doublets homonymiques</u>	<u>Nombre d'erreurs</u>
1. é ; er	20
2. à ; a	12
3. qui ; qu'il	9
4. s'est ; c'est ces ; ses sais ; sait	9
5. ce ; se	7
6. on ; ont	5
7. sont ; son	2
8. dont ; donc	2
9. ait ; est	2
10. où ; ou	2
11. près ; prêt	2
12. quelle ; qu'elle	1
13. soi ; soit	1
14. d'en ; dans	1
15. sans ; s'en	1
16. là ; la	1
17. tant ; temps	1
Total	<u>77</u>

Erreurs non-attribuables aux conflits homonymiques

Total	97
-------	----

TABLEAU B.9
Détail des erreurs d'usage lexical

<u>Type d'erreur</u>	<u>Origine de l'erreur</u>		<u>Pourcentage</u>
	Interférence de l'anglais	Non interférence	
Prépositions	32	47	33.3
Verbes	26	45	29.9
Substantifs	16	31	19.8
Adjectifs	7	16	" 9.7
Pronoms	-	8	3.3
Adverbes	-	6	2.5
Article	-	1	0.4



APPENDIX C

Basis for Classifying University Courses
(for Classification Codes, see end of Table)

P	-	Administration	P	-	Hotel and Food Administration
P	-	Agriculture	P	-	Industrial Design
SS	-	Amerindian Eskimo Studies	SS	-	International Studies
S	-	Anatomy	H	-	Italian
P	-	Animal Science	P	-	Journalism
SS	-	Anthropology	SS	-	Latin American Studies
P	-	Applied Computational and Math Science	P	-	Law
P	-	Applied Human Nutrition	S	-	Liberal Sciences
P	-	Applied Statistics and Computation	P	-	Library Technology
P	-	Architecture	S	-	Life Sciences
SS	-	Asian Studies	H	-	Linguistics
S	-	Astronomy	S	-	Marine Biology
S	-	Biochemistry	S	-	Mathematics
S	-	Biology	P	-	Medical Lab Sciences
S	-	Biophysics	P	-	Medicine
S	-	Botany	H	-	Medieval Studies
P	-	Business Administration	P	-	Metallurgy
SS	-	Canadian Studies	S	-	Microbiology
S	-	Chemistry	H	-	Music
SS	-	Child Studies	SS	-	Native Studies
SS	-	Classical Archaeology	P	-	Nursing
H	-	Classics (Latin and Greek)	P	-	Optometry
P	-	Commerce	P	-	Pharmacy
SS	-	Commonwealth Studies	H	-	Philosophy
P	-	Communication Arts	P	-	Physical Education and Kinesiology
H	-	Comparative Literature	S	-	Physical Geography
P	-	Computer Science	S	-	Physics
SS	-	Consumer Studies	P	-	Physiology

P	-	Dental Hygiene	S	-	Plant Science
P	-	Dentistry	SS	-	Political Science
SS	-	Development Studies	S	-	Probabilities and Statistics
H	-	Drama	SS	-	Psychology
S	-	Earth Sciences	P	-	Public Administration
SS	-	Economics	P	-	Recreational Studies
P	-	Education	P	-	Regional Environmental Planning
P	-	Engineering	P	-	Rehabilitative Medicine
P	-	Engineering and Management	H	-	Religion/Religious Studies
H	-	English	S	-	Science, General
S	-	Environmental Science	SS	-	Slavik Studies
SS	-	Family Studies	SS	-	Social Theory Program
H	-	Film	P	-	Social Work
H	-	Fine Arts/Art	SS	-	Sociology
P	-	Fisheries and Wildlife Biology	SS	-	Soviet and East European Studies
P	-	Forestry	H	-	Spanish
H	-	French	P	-	Sports Administration
P	-	Fr. Language Translation and Studies	S	-	Statistics
S	-	Genetics	SS	-	Studies in Social Action
SS	-	Geography	P	-	Survey Science
S	-	Geology	P	-	Textiles
H	-	German	P	-	Theology
P	-	Health	P	-	Urban Studies
SS	-	History	P	-	Veterinary Medicine
P	-	Home Economics	S	-	Zoology

Note: The four program areas and corresponding letter codes are as follows:

H - Humanities/Arts
 SS - Social Sciences
 S - Sciences
 P - Professions
 i.e. Applied Arts
 Applied Social Sciences
 Applied Sciences

APPENDIX D1

EQUATING SCORES ON BOTH FORMS OF THE TEST OF READING COMPREHENSION AND LANGUAGE ACHIEVEMENT (ENGLISH) AND THE TEST DE COMPREHENSION EN LECTURE ET DE CONNAISSANCE DE LA LANGUE (FRANCAIS)

1. PROCEDURE

The rationale for this equating procedure was proposed by Lord (1950).
The procedure itself is described by Angoff (1971).

1.1 Required Data

Each of a group of examinees takes both forms of the test. The following subscripts are used:

$i = 1, \dots, n_j$ examinees

$j = a, b$ groups of examinees

$k = 1, 2$ test forms

The n_a examinees in group "a" take the forms in the order Form 1, Form 2; the n_b examinees in group "b" take the forms in the order Form 2, Form 1. The scores are denoted as: x_{kij} = mark achieved on form k by examinee i of group j.

1.2 Equations

The basic equation is:

$$\frac{X_{1i} - M_1}{s_1} = \frac{X_{2i} - M_2}{s_2}$$

where

M_k is the mean of form k

s_k is the standard deviation of form k

and these parameters are estimated as follows:

$$M_k = .5 (M_{ka} + M_{kb} - C_k)$$

$$s_k^2 = .5 (s_{ka}^2 r_{kka} + s_{kb}^2 r_{k kb})$$

$$\frac{C_k}{s_k} = H$$

$$H = .5 \left(\frac{M_{1b} - M_{1a}}{s_1} + \frac{M_{2a} - M_{2b}}{s_2} \right)$$

The remaining undefined terms r_{kka} and $r_{k kb}$ are the reliability estimates of form k from the scores obtained by groups á and b respectively.

Given estimates of the parameters in the basic equation, scores from one form of the test can be taken into the metric of the other form of the test.

2. APPLICATION

The data for equating scores on the forms of the Test of Reading Comprehension and Language Achievement (English) are presented in Table D1.1. The corresponding data for the Test de compréhension en lecture et de connaissance de la langue (français) are given in Table D1.2.

It was decided, arbitrarily, to equate scores to the form of the test having the higher mean. For the English test, Form 2 scores were taken into the metric of Form 1. For the français test, Form 1 scores were taken into the metric of Form 2.

2.1 Equations for the Anglophone Test

(i) Reading Comprehension Subtest:

$$\hat{X}_1 = 1.10X_2 + 0.30$$

(ii) First Language Achievement Subtest:

$$\hat{X}_1 = 1.03X_2 - 0.26$$

(iii) Second Language Achievement Subtest:

$$\hat{X}_1 = 0.97X_2 + 1.58$$

(iv) Total Test:

$$\hat{X}_1 = 1.01X_2 + 1.69$$

2.2 Equations for the Francophone Test

(i) Reading Comprehension Subtest:

$$\hat{X}_2 = 1.02X_1 + 0.99$$

(ii) First Language Achievement Subtest:

$$\hat{X}_2 = 1.01X_1 + 0.29$$

(iii) Second Language Achievement Subtest:

$$\hat{X}_2 = 1.21X_1 + 1.04$$

(iv) Total Test:

$$\hat{X}_2 = 1.08X_1 + 2.12$$

In the foregoing equations, the subscripts 1 and 2 refer to test form, X refers to test score, and the symbol " $\hat{}$ " over X designates the fact that this is a score estimated from the student's observed score on the other form.

TABLE D1.1

Data for Equating Both Forms of the Test of Reading Comprehension and Language Achievement (English)

<u>Test Part</u>	<u>Form</u>	<u>Statistic</u>	<u>Group 1^a</u> <u>(N=1612)</u>	<u>Group 2^a</u> <u>(N=1619)</u>
Reading Comprehension Subtest	1	Mean	4.63	4.67
		s.d. ^b	2.61	2.69
		rx _x	0.67	0.68
	2	Mean	4.01	3.93
		s.d.	2.59	2.58
		rx _x	0.59	0.60
First Language Achievement Subtest	1	Mean	4.20	4.78
		s.d.	2.95	2.98
		rx _x	0.66	0.64
	2	Mean	5.05	4.19
		s.d.	2.78	2.85
		rx _x	0.67	0.70
Second Language Achievement Subtest	1	Mean	7.66	7.99
		s.d.	3.30	3.27
		rx _x	0.61	0.62
	2	Mean	6.58	6.33
		s.d.	3.43	3.40
		rx _x	0.61	0.61
Total Test	1	Mean	16.49	17.44
		s.d.	7.45	7.53
		rx _x	0.83	0.83
	2	Mean	15.64	14.45
		s.d.	7.40	7.35
		rx _x	0.83	0.83

^aGroup 1 took the forms in the order Form 1 - Form 2; Group 2 took them in reverse order.

^brx_x symbolizes the coefficient of reliability, estimated using Kuder-Richardson formula 20 (Lord & Novick, 1968, p.91).

TABLE D1.2

Equating Data for Both Forms of the Test de compréhension en lecture et de connaissance de la langue (français)

<u>Test Part</u>	<u>Form</u>	<u>Statistic</u>	<u>Group 1^a</u> <u>(N=248)</u>	<u>Group 2^a</u> <u>(N=249)</u>
Reading Comprehension Subtest	1	Mean	5.12	5.09
		s.d. ^b	2.83	3.08
		rxx	0.59	0.66
	2	Mean	6.24	6.13
		s.d.	2.88	3.30
		rxx	0.52	0.64
First Language Achievement Subtest	1	Mean	2.40	2.28
		s.d.	2.58	2.70
		rxx	0.52	0.55
	2	Mean	2.69	2.63
		s.d.	2.65	2.66
		rxx	0.54	0.54
Second Language Achievement Subtest	1	Mean	1.70	2.28
		s.d.	2.37	2.66
		rxx	0.53	0.59
	2	Mean	3.57	3.33
		s.d.	2.78	2.80
		rxx	0.66	0.68
Total Test	1	Mean	9.23	9.64
		s.d.	6.21	6.66
		rxx	0.76	0.78
	2	Mean	12.50	12.09
		s.d.	6.63	7.10
		rxx	0.77	0.80

^aGroup 1 took the forms in the order Form 1-Form 2; Group 2 took them in reverse order.

^brxx symbolizes the coefficient of reliability, estimated using Kuder-Richardson formula 20 (Lord & Novick, 1968, p.91).

APPENDIX D2

PROCEDURE FOR SCORING ESSAYS

ANALYSIS OF THE CALIBRATION SAMPLE

Data. Each of m scorers marked each of n essays. The following indexes will be used:

$$i = 1, \dots, n \text{ essays}$$

$$p = 1, \dots, m \text{ scorers.}$$

The marks will be denoted as:

$$y_{ip} = \text{mark given by scorer } p \text{ to essay } i.$$

Model. It will be assumed that there is a single common characteristic being measured. The scorers vary in their mean levels of scoring, their scales, and their reliabilities. The fundamental equation is:

$$y_{ip} = a_p + b_p t_i + e_{ip}$$

where

$$t_i = \text{latent, common characteristic for essay } i,$$

$$a_p = \text{mean score level for scorer } p,$$

$$b_p = \text{scaling factor for scorer } p,$$

$$e_{ip} = \text{"error"}.$$

The error terms are assumed to be independent within and across scorers and essays. The errors are distributed normally and homogeneously by scorer.

$$e_{ip} \sim N(0, V_p)$$

where

$$V_p = \text{error (unique) variance for scorer } p.$$

Identification. To define a unique solution, we will fix the mean and variance of the t_i .

$$E(t) = 0$$

$$V(t) = 1$$

(This may later be changed to any desired scale.)

Estimation. Using the matrix $y = (y_{ip})$ we perform a 1-dimensional common factor analysis. This involves computing:

$$\bar{y}_p = \text{mean mark for scorer } p,$$

$$s_p^2 = \text{variance of marks for scorer } p,$$

$$r_{pq} = \text{correlation of marks for scorers } p \text{ and } q,$$

and gives as results:

$$h_p^2 = \text{estimated communality for scorer } p,$$

$$h_p = \text{factor loading for scorer } p,$$

$$1-h_p^2 = \text{unique variance for scorer } p.$$

Transformation. The factor analysis results can be translated into estimates of the original model parameters:

$$\hat{a}_p = \bar{y}_p,$$

$$\hat{b}_p = s_p h_p,$$

$$\hat{v}_p = s_p^2 (1-h_p^2).$$

SCORING THE DATA

Data. Suppose that for a particular essay we have a (reduced) set of scorings:

$$y_p, p = 1, \dots, m^*.$$

We want to estimate t for the essay.

Model. Following the model for the calibration data, we assume that

$$y_p \sim N(a_p + b_p t, V_p), \quad p = 1, \dots, m^*,$$

and that the "errors" are independent.

Derivation. The log-likelihood of the observed set of scores is

$$\ell = -\frac{1}{2} \sum_p \frac{(y_p - a_p - b_p t)^2}{\sqrt{V_p}} + \text{constant}.$$

We differentiate ℓ with respect to t and set to zero:

$$\frac{d\ell}{dt} = \sum_p \frac{b_p}{\sqrt{V_p}} \frac{(y_p - a_p - b_p t)}{\sqrt{V_p}} = 0.$$

The maximum-likelihood solution for t is

$$\hat{t} = \frac{\sum_p \frac{b_p}{V_p} (y_p - a_p)}{b_p^2 \sum_p \frac{1}{V_p}}.$$

Substitution. For practical calculations, we substitute the estimates of the parameters:

$$\hat{t} = \frac{\sum_p \frac{s_p^2 h_p^2}{s_p^2 (1-h_p^2)} (y_p - \bar{y}_p)}{\sum_p \frac{s_p^2 h_p^2}{s_p^2 (1-h_p^2)}}$$

$$= \frac{\sum_p r_p \frac{(y_p - \bar{y}_p)}{s_p h_p}}{\sum_p r_p}$$

where

$$r_p = \frac{h_p^2}{1-h_p^2}$$

is the signal to noise ratio for scorer p and has the effect of emphasizing the more reliable scores. Note also that the reliabilities are weighting the estimates as follows:

$$\frac{y_p - \bar{y}_p}{s_p h_p}$$

so h_p has the effect of expanding the scores of the less reliable scorers.

IMPLEMENTATION

The essay scoring procedure was applied to four sets of essay marks: those for the Writing Test (English); Test de composition écrite (français); Writing Exercise (anglais) - Score on the Summary; and Writing Exercise (anglais) - Score on the Commentary. The data required to implement the procedure include, for each marker, the mean and standard deviation of scores assigned to essays in the calibration sample and the factor loading of the marker on the first common factor of the matrix of intercorrelations among markers. (The correlations in this matrix are compiled using the essays in the calibration sample as observations.)

The data used to implement the essay scoring procedure for each of the four sets of essay marks are reported in Table D2.1. Given these figures, the raw scores assigned to an essay, and the identification numbers of the individuals who marked the essay, the equations of the preceding section can be used to derive the score for the essay. This score can be regarded as a weighted average of raw scores, where the weights make the necessary adjustment to allow for differences in the mean, the variability and the consistency of the marks assigned by different markers.

After all the essays had been scored, a final linear adjustment was made to ensure that the distribution of scores on the essays exactly covered the full scale of marks that the scorers had been told to use. In the cases of the Writing Test (English), the Test de composition écrite (français) and the Writing Exercise (anglais) - Score in the Commentary, this scale of marks was 1 to 10. The scale of marks on the Writing Exercise (anglais) - Score on the Summary ran from 0 to 10.

TABLE D2.1

Data Required to Implement the Essay
Scoring Procedure

<u>Scorer</u>	<u>Mean</u>	<u>S.D.</u>	<u>Factor Loading</u>
Writing Test (English)			
1	5.61	1.86	0.72
2	5.25	1.87	0.66
3	5.86	1.64	0.81
4	5.60	1.85	0.91
5	5.30	2.43	0.72
6	5.60	2.52	0.76
7	4.91	1.83	0.70
8	4.72	2.74	0.66
9	6.00	2.21	0.72
10	5.23	2.30	0.86
11	4.45	2.05	0.83
12	5.64	2.60	0.76
13	5.21	2.46	0.68
14	5.24	2.45	0.65
15	5.81	2.58	0.80
16	5.34	1.78	0.81
17	6.42	2.24	0.73
18	5.92	2.00	0.68
19	6.02	2.48	0.73
20	4.69	2.37	0.79
21	5.35	2.50	0.87
22	4.98	2.33	0.60
23	4.72	1.94	0.59
24	5.15	2.32	0.88
25	6.54	2.35	0.67
26	5.75	2.21	0.79
27	5.75	1.80	0.54
28	5.42	2.29	0.74
29	5.41	2.10	0.76
30	5.98	1.91	0.72
31	5.73	2.15	0.71
32	6.26	1.71	0.86
33	5.52	1.90	0.85
34	6.25	1.47	0.72
35	5.60	2.29	0.74
36	6.19	2.64	0.76

TABLE D2.1 (continued)

<u>Scorer</u>	<u>Mean</u>	<u>S.D.</u>	<u>Factor Loading</u>
---------------	-------------	-------------	-----------------------

Composition écrite (français)

1	5.98	1.18	0.63
2	5.71	1.71	0.78
3	5.43	2.43	0.78
4	6.20	1.97	0.86
5	6.20	2.42	0.83
6	6.50	1.81	0.89
7	6.65	2.29	0.68
8	5.26	1.63	0.52
9	6.08	1.56	0.78

Writing Exercise (anglais) - Score on Summary

1	3.05	2.21	0.46
2	5.80	2.50	0.78
3	2.45	2.52	0.70
4	4.45	1.85	0.74
5	2.90	1.56	0.91
6	3.20	1.88	0.88

Writing Exercise (anglais) - Score on Commentary

1	3.10	2.29	0.84
2	3.65	3.30	0.94
3	2.05	2.54	0.92
4	3.40	2.93	0.78
5	2.65	2.13	0.91
6	2.45	2.91	0.95

APPENDIX E
STUDENT QUESTIONNAIRE

Instructions

In order for this study to be carried out properly, it is important for us to have certain information about you, your family, and your plans for the future. This information will be kept completely confidential. It will be read only by the person responsible for computer coding your responses. When the information has been coded and stored in the computer, it will be identified with a code number. The list which matches code numbers with individual students will be accessible to only a small handful of researchers.

There are some special instructions for completing the questionnaire. Please read these instructions carefully before you begin.

1. NUMERICAL ANSWERS. In some cases you will be asked to give your answer in numbers in the right-hand column of the page. If you are giving an answer in numbers, please use enough digits in your answer to fill all the spaces provided. This will mean that in some cases you should put zeros before your answer. For example, if your answer is "3", and two spaces are provided, you should write your answer as "03".

2. VERBAL ANSWERS. In other cases, you will be asked to give your answer in words in a space provided at the left side of the page. You will notice that in the right-hand column opposite each of these spaces there are parentheses. DO NOT WRITE ANYTHING IN THESE PARENTHESES. They are for use in translating your verbal answer into a number to be coded.

Please try to answer all questions as completely and accurately as possible. Any omitted or unclear answers affect the accuracy of the study results.

Thank you for your help.

Student's Name _____

SECTION 1: PERSONAL INFORMATION

1A. What is your date of birth?

Day (2-digit number: 01,
02, ...)

Month (2-digit number:
Jan = 01, Feb = 02, ...)

Year

1B. Write the following number:

1 if **you** are female
2 if **you** are male

SECTION 2: LANGUAGE

2A. What language do your parents or guardians usually speak in your home? Write the number corresponding to this language in the list below.

- 1: French
- 2: English
- 3: Other

If you answered "Other" by writing the number 3, please write the name of the language here:

2B. What language do you usually speak outside school and home, with your friends and neighbours? Write the number corresponding to this language in the list below.

- 1: French
- 2: English
- 3: Other

If you answered "Other" by writing the number 3, please write the name of the language here:

Questions 2C, 2D, and 2E refer to full-time regular school attendance, from kindergarten to your present grade. Do not include information about special schools attended outside regular school hours (e.g., on Saturdays, or on weekday afternoons after your regular school day).

2C. How many years (counting the present year, if applicable) have you spent in schools in which most or all of your classes were taught in French? (Give as a 2-digit number: 00, 01, 02, ...)

2D. How many years (counting the present year, if applicable) have you spent in schools in which most or all of your classes were taught in English? (Give as a 2-digit number: 00, 01, 02, ...)

- 2E. How many years have you spent in schools in which most
or all of your classes were taught in a language other
than French or English? (Give as a 2-digit number:
00, 01, 02, ...)

Write the name(s) of the language(s) here:

SECTION 3: COUNTRY OF BIRTH

3A. In what country were you born? Write the number corresponding to the country in the list below.

- 1: Canada
- 2: United States
- 3: Great Britain
- 4: France
- 5: Other

If you answered "Other" by writing the number 5, please write the name of the country here:

ANSWER QUESTION 3B ONLY IF YOU WERE NOT BORN IN CANADA. IF YOU WERE BORN IN CANADA, LEAVE QUESTION 3B BLANK AND GO DIRECTLY TO SECTION 4.

3B. In what year did you come to Canada?

In the next three sections of the questionnaire (Sections 4, 5, and 6), you will be asked for information about your parents and/or guardians. We realize that in many cases the adults who are responsible for you are people other than your natural parents. Therefore, each of these questions has three parts:

- a) Information about your natural parents. If you are able to, please give this information whether or not you actually live with one or both natural parents.
- b) Information about your adoptive parents. If you have been legally adopted, please give this information whether or not you actually live with one or both adoptive parents.
- c) Information about your guardians. Please give this information if your present male or female guardian is someone other than a natural or adoptive parent. For example, you may have a stepfather or stepmother, or may live with grandparents or in a foster home.

In these sections, leave an answer space blank only if the particular question is not applicable to your situation. For example, if you have an adoptive father, do not leave blank spaces for the questions relating to adoptive fathers, even if you do not have the required information. Instead, fill in the code number which matches the answer "Unknown".

SECTION 4: PARENTS' AND/OR GUARDIANS' COUNTRIES OF BIRTH

BEFORE ANSWERING THE QUESTIONS IN THIS SECTION, PLEASE REFER TO THE INSTRUCTIONS ON THE BACK COVER FOR GUIDANCE.

ANSWER THESE QUESTIONS BY WRITING THE NUMBER CORRESPONDING TO THE CORRECT ANSWER IN THIS LIST:

- 1: Canada
- 2: United States
- 3: Great Britain
- 4: France
- 5: Other
- 6: Unknown

4A. In what country was your natural father born?

If you answered "Other" by writing the number 5, please write the name of the country here:

4B. In what country was your natural mother born?

If you answered "Other" by writing the number 5, please write the name of the country here:

4C. (Answer if you have an adoptive father) In what country was your adoptive father born?

If you answered "Other" by writing the number 5, please write the name of the country here:

4D. (Answer if you have an adoptive mother) In what country was your adoptive mother born?

If you answered "Other" by writing the number 5, please write the name of the country here:

- 4E. (Answer if your present male guardian is someone other than your natural or adoptive father) In what country was your present male guardian born?

If you answered "Other" by writing the number 5, please write the name of the country here.

- 4F. (Answer if your present female guardian is someone other than your natural or adoptive mother) In what country was your present female guardian born?

If you answered "Other" by writing the number 5, please write the name of the country here:

SECTION 5: EDUCATION OF PARENTS AND/OR GUARDIANS

BEFORE ANSWERING THE QUESTIONS IN THIS SECTION, PLEASE REFER TO THE INSTRUCTIONS ON THE BACK COVER FOR GUIDANCE.

ANSWER THESE QUESTIONS BY WRITING THE NUMBER CORRESPONDING TO THE CORRECT ANSWER IN THIS LIST. IF YOU ARE UNCERTAIN OF THE CORRECT ANSWER, BUT HAVE SOME REASONABLY GOOD IDEA, PLEASE PICK THE ANSWER YOU BELIEVE MOST LIKELY. ANSWER "UNKNOWN" ONLY IF YOU HAVE NO KNOWLEDGE AT ALL OF THE CORRECT ANSWER.

- 1: Unknown.
- 2: Did not complete elementary school.
- 3: Completed elementary school, but did not continue to secondary school.
- 4: Attended secondary school, but did not graduate.
- 5: Graduated from secondary school, but did not continue to a postsecondary educational institution (e.g., university, community college, art college, agricultural college).
- 6: Attended a postsecondary educational institution, but did not graduate.
- 7: Graduated from a postsecondary educational institution.

5A. Which statement most accurately describes the highest level of your natural father's education?

5B. Which statement most accurately describes the highest level of your natural mother's education?

5C. (Answer if you have an adoptive father) Which statement most accurately describes the highest level of your adoptive father's education?

5D. (Answer if you have an adoptive mother) Which statement most accurately describes the highest level of your adoptive mother's education?

5E. (Answer if your present male guardian is someone other than your natural or adoptive father) Which statement most accurately describes the highest level of your present male guardian's education?

5F. (Answer if your present female guardian is someone other than your natural or adoptive mother) Which statement most accurately describes the highest level of your present female guardian's education?

SECTION 6: OCCUPATIONS OF PARENTS AND/OR GUARDIANS

BEFORE ANSWERING THE QUESTIONS IN THIS SECTION, PLEASE REFER TO THE INSTRUCTIONS ON THE BACK COVER FOR GUIDANCE.

You will be asked below to give the occupations of your parents and/or guardians. Please read these instructions carefully before answering.

If you do not know the occupation of one of your parents or guardians, write the word "unknown".

If possible, give a two-word description of the occupation--e.g., "bus driver", "industrial accountant", "telephone operator", "dairy farmer", "university student".

Try to make your description as specific as possible. For example, an answer of "restaurant business" is much less informative than one of "restaurant waiter" or "restaurant owner". You need not include the name of the employer. For example, write "sales clerk" rather than "Eaton's clerk".

If one of your parents or guardians has more than one occupation, list only the occupation at which he or she spends the most time.

If the main occupation of one of your parents or guardians is looking after the home, list his or her occupation as "homemaker".

If one of your parents or guardians is deceased, retired, or unemployed, write "deceased", "retired", or "unemployed", followed by the name of his or her most recent occupation.

6A. What is the occupation of your natural father?

6B. What is the occupation of your natural mother?

6C. (Answer if you have an adoptive father) What is the occupation of your adoptive father?

6D. (Answer if you have an adoptive mother) What is the occupation of your adoptive mother?

6E. (Answer if your present male guardian is someone other than your natural or adoptive father) What is the occupation of your present male guardian?

6F. (Answer if your present female guardian is someone other than your natural or adoptive mother) What is the occupation of your present female guardian?

SECTION 7: EDUCATIONAL PLANS (I)

The term "postsecondary educational institution" refers to any institution normally attended after leaving secondary school - for example, a university, a community college, an art college, or an agricultural college.

7A. Suppose that in Ontario all postsecondary educational programs now available in English were also available in French. In what way, if any, would this affect your present plans for postsecondary education? Choose the most appropriate statement below, and write the corresponding number in the space.

- 1: There would be no change in my plans.
- 2: I do not now plan to pursue postsecondary education. I would do so if an appropriate course were available in French.
- 3: I now plan to pursue postsecondary education in English. I would pursue the same course in French if it were available.
- 4: I now plan to pursue postsecondary education in English. I would pursue a different course in French if it were available.
- 5: I now plan to pursue postsecondary education in French. I would pursue a different course in French if it were available.

7B. Beginning in September 1976, which of the following do you expect to be doing? Write the number that corresponds to your answer in the list.

- 1: Attending secondary school.
- 2: Attending a postsecondary educational institution.
- 3: Working full time.
- 4: Other.

If you answered "Other" be writing the number 4, please explain here what you expect to be doing:

IF YOU ANSWERED QUESTION 7B BY STATING THAT YOUR PLANS FOR SEPTEMBER 1976 ARE "ATTENDING SECONDARY SCHOOL", GO ON TO SECTION 8.

IF YOU ANSWERED QUESTION 7B BY STATING THAT YOUR PLANS FOR SEPTEMBER 1976 ARE "ATTENDING A POSTSECONDARY EDUCATIONAL INSTITUTION", GO ON TO SECTION 9.

IF YOU ANSWERED QUESTION 7B BY STATING THAT YOUR PLANS FOR SEPTEMBER 1976 ARE "WORKING FULL TIME" OR "OTHER", GO ON TO SECTION 10.

SECTION 8: EDUCATIONAL PLANS (II)

ANSWER THE QUESTIONS IN THIS SECTION ONLY IF YOU ANSWERED QUESTION 7B BY STATING THAT YOUR PLANS FOR SEPTEMBER 1976 ARE "ATTENDING SECONDARY SCHOOL".

8A. Do you expect to begin attending a postsecondary educational institution at some time before the end of 1979? Write 1 for "yes", 2 for "no".

8B. (Answer if your answer to Question 8A was "yes")
About how many years do you expect to spend studying at the postsecondary level? (Give as a 2-digit number: 01, 02, ...)

GO ON TO SECTION 11.

SECTION 9: EDUCATIONAL PLANS (III)

ANSWER THE QUESTIONS IN THIS SECTION ONLY IF YOU ANSWERED QUESTION 7B BY STATING THAT YOUR PLANS FOR SEPTEMBER 1976 ARE "ATTENDING A POSTSECONDARY EDUCATIONAL INSTITUTION".

- 9A. About how many years do you expect to spend studying at the postsecondary level? (Give as a 2-digit number: 01, 02, ...)
- 9B. Have you applied for 1976 admission to one or more postsecondary educational institutions? Write 1 for "yes", 2 for "no".
- 9C. (Answer if your answer to Question 9B was "yes") List below, in order of your preference, the institutions and programs to which you have applied for 1976 admission. For each, please give the following information:

Name of institution (e.g., University of Toronto, Cambrian College)

Program applied for (e.g., Biology, Engineering, Early Childhood Education)

Language in which the program is taught (French or English)

Degree, diploma, certificate, etc., which you hope to attain (e.g., B.A., Diploma in Agriculture)

1. Institution _____
Program _____
Language _____
Degree, etc. _____
2. Institution _____
Program _____
Language _____
Degree, etc. _____

(list continues on next page)

3. Institution _____
Program _____
Language _____
Degree, etc. _____
4. Institution _____
Program _____
Language _____
Degree, etc. _____
5. Institution _____
Program _____
Language _____
Degree, etc. _____

9D. Unless you give permission for their release, your marks on any tests you write for this project will be given only to you. However, you may choose to have your marks given to your postsecondary institution for their research purposes, for their use in counselling you, or for both. If these marks are released to such an institution with your consent, they will not be used in making any decision about your admission to the institution; in fact, in most cases the institution would not receive the marks until after a decision on your admission had already been made.

Are you willing to have your score(s) on any test(s) you write for this project made available to your postsecondary institution:

- a) for research purposes? Write 1 for "yes", 2 for "no".
- b) for use in counselling you? Write 1 for "yes", 2 for "no".

GO ON TO SECTION 11.

SECTION 10: EMPLOYMENT PLANS

ANSWER THE QUESTIONS IN THIS SECTION ONLY IF YOU ANSWERED QUESTION 7B BY STATING THAT YOUR PLANS FOR SEPTEMBER 1976 ARE "WORKING FULL TIME".

- 10A. What job do you expect to be working at in September 1976? If possible, give a two-word description.

- 10B. Do you have a job offer for 1976? Write 1 for "yes, 2 for "no".

If you answered "yes" by writing the number 1, what is the job? If possible, give a two-word description.

SECTION 11: CAREER PLANS

- 11A. Have you made a decision about what your eventual career will be? Write 1 for "yes", 2 for "no".

If you answered "yes" by writing the number 1, what is the career? If possible, give a two-word description.

APPENDIX F
TABLES OF STATISTICS REFERRED TO IN THE
SUBSECTION ON MARKING STANDARDS (CHAPTER 3, 1.3)
OF THE SSGD/SSHGD SURVEY
OF ANGLOPHONE STUDENTS

TABLE F.1

Summary Statistics From the Multiple-Regression Analyses of
Marks in English Courses

Variable ^a	S.D.	B	S.E. of B	Intercorrelations			
				1	2	3	4
SSHGD Analysis ^a							
1. School Marks	9.7			-			
2. Reading Comprehension Part	2.4	0.70	0.18	0.41	-		
3. First Language Achievement Part	2.6	0.60	0.16	0.41	0.43	-	
4. Second Language Achievement Part	2.8	0.51	0.16	0.41	0.45	0.49	-
5. <u>Writing Test</u>	1.1	0.27	0.04	0.46	0.33	0.33	0.33
Multiple Correlation Coefficient = 0.58; N=483; n=50							
SSGD Analysis - Advanced Level Courses							
1. School Marks	10.6						
2. Reading Comprehension Part	2.5	0.43	0.21	0.33	-		
3. First Language Achievement Part	2.5	0.67	0.21	0.36	0.47	-	
4. Second Language Achievement Part	2.8	0.42	0.18	0.32	0.44	0.42	-
5. <u>Writing Test</u>	1.1	0.24	0.04	0.41	0.40	0.40	0.37
Multiple Correlation Coefficient = 0.49; N=450; n=47							

TABLE F.1 (continued)

Variable ^a	S.D.	B	S.E. of B	Intercorrelations			
				1	2	3	4
SSHGD Analysis ^a							
SSGD Analysis - General Level Courses							
1. School Marks	9.0			-			
2. Reading Comprehension	2.0	0.55	0.30	0.29	-		
3. First Language Achievement Part	2.1	0.58	0.29	0.32	0.32	-	
4. Second Language Achievement Part	2.5	0.68	0.24	0.36	0.24	0.32	-
5. Writing Test	1.0	0.23	0.06	0.41	0.33	0.35	0.36

Multiple Correlation Coefficient = 0.49; N=220; n=46

Note: Statistics S.D., B--the raw score "partial" regression coefficient, S.E. of B, and the interrelations among the variables were derived from pooled within-school variation and covariation. N is the number of students, n the number of secondary schools.

^aVariables 2, 3 and 4 were scores on the constituent parts of the Test of Reading Comprehension and Language Achievement (English). Students with marks in more than one English course were assigned an average mark.

TABLE F.2

Summary Statistics From the Multiple-Regression Analysis
of Marks in SSHGD French Courses on Scores in the
French Reading Test and the French Listening Test

<u>Variable</u>	<u>S.D.</u>	<u>B</u>	<u>S.E. of B</u>	Intercorrelations	
				<u>1</u>	<u>2</u>
1. <u>School Marks^a</u>	9.1			-	
2. <u>French Reading Test</u>	6.5	0.71	0.09	0.67	-
3. <u>French Listening Test</u>	6.8	0.31	0.08	0.58	0.70

Multiple Correlation Coefficient = 0.69; N=270; n=47

Note: Statistics S.D., B--the raw score "partial" coefficient of regression, S.E. of B, and the intercorrelations among the variables were derived from pooled within-school variation and covariation. N is the number of students, n the number of secondary schools.

^aStudents earning marks in more than one SSHGD level French course were assigned an average rank.

TABLE F.3

Summary Statistics From the Regression
Analyses of Marks in Mathematics and Courses

<u>Variable</u>	<u>S.D.</u>	<u>B</u>	<u>S.E. of B</u>	<u>r</u>	<u>N</u>	<u>n</u>
SSHGD Analysis						
1. School Marks ^a in Mathematics	12.6					
2. <u>Mathematics</u> <u>Achievement Test</u>	5.6	1.47	0.06	0.65	842	51
1. School Marks in Physics	11.7					
2. <u>Physics</u> <u>Achievement Test</u>	8.3	0.84	0.06	0.60	376	52
SSGD Analysis						
1. School Marks in Advanced Mathematics ^b	13.5					
2. <u>Test of Arithmetic</u> <u>and Basic Algebra</u>	6.1	1.35	0.07	0.61	684	46
1. School Marks in General Mathematics ^c	11.8					
2. <u>Arithmetic</u> <u>and Basic Algebra</u>	5.5	1.00	0.09	0.47	420	43

Note: Statistics S.D., B--the raw score regression coefficient, S.E. of B, and r--the correlation coefficient, were derived from pooled within-school variation and covariation. N is the number of students, n the number of secondary schools.

^a Average of marks in the calculus course and the functions and relations course.

^b Mark in the Grade 12 "foundations" mathematics course.

^c Mark in the Grade 12 "applications" mathematics course.

TABLE F.4

Means, Standard Deviations and Sample Sizes for
Distribution of Achievement Test Scores and School Marks
of SSHGD Students

<u>Variable</u>	<u>N</u> <u>(3410)^a</u>	<u>Mean</u>	<u>S.D.</u>
<u>Test of Reading Comprehension and</u> <u>Language Achievement (English)</u>			
1. Reading Comprehension Part (RC)	2255	5.1	2.6
2. First Language Achievement Part (L1)	2255	4.4	2.9
3. Second Language Achievement Part (L2)	2255	8.1	3.2
4. Total Test (RCLA)	2255	17.5	7.0
<u>Writing Test (W)</u>	569	6.4	1.3
<u>Tests of French as a Second Language</u>			
1. <u>Reading Test (FR)</u>	360	22.2	7.6
2. <u>Listening Test (FL)</u>	288	19.1	8.9
<u>Mathematics Achievement Test (MAT)</u>	885	17.2	6.0
<u>Physics Achievement Test (PAT)</u>	401	15.6	9.7
<u>School Marks</u>			
1. English (E)	2678	68.0	11.2
2. French (F)	690	71.5	11.5
3. Mathematics (M)	2177	68.6	14.5
4. Physics (P)	1168	68.4	13.3

^aThis was the number of students drawn in the original sample. The difference between this number and the number of students who took the Test of Reading Comprehension and Language Achievement is accounted for by a small number of unusable test records and absenteeism on the day of the test.

TABLE F.5

Correlation Coefficients and Sample Sizes for Correlation Coefficients
Among Achievement Tests and School Marks for SSHGD Students

Variables ^a	RC	L1	L2	RCLA	W	FR	FL	MAT	PAT	E	F	M	P
RC		2555	2555	2555	566	340	271	854	398	1865	506	1557	840
L1	48		2555	2555	566	340	271	854	398	1865	506	1557	840
L2	49	51		2555	506	340	271	854	398	1865	506	1557	840
RCLA	78	81	85		506	340	271	854	398	1865	506	1557	840
W	38	36	39	46		90	67	202	^b	486	133	371	139
FR	33	34	33	41	24		285	-	^c	336	341	195	54
FL	18	20	20	24	26	73		-	X	273	273	159	45
MAT	31	32	26	36	27	-	-		253	701	90	842	522
PAT	36	26	22	34	-	X	X	66		299	48	363	376
E	41	40	40	50	42	36	27	33	20		622	1729	854
F	33	36	37	43	37	62	53	23	08	56		390	126
M	25	24	23	29	18	35	21	60	38	50	55		1056
P	28	28	26	33	15	21	22	58	54	56	54	73	

TABLE F.5 (Continued)

Note: Correlation coefficients appear below the diagonal; the size of the sample for each coefficient appears above the diagonal. Decimal points have been omitted from the correlation coefficients.

^aFor a key to the variable labels, see Table F.4

^bCells marked with a dash involve combinations of tests taken by no student.

^cCells marked with an "X" involved fewer than 30 students; results are not reported for these combinations of variables.

TABLE F.6
Means, Standard Deviations and Sample Scores for Distributions
of Achievement Test Scores and School Marks Obtained on
SSGD Students

<u>Variable</u>	<u>N</u> <u>(2558)^a</u>	<u>Mean</u>	<u>S.D.</u>
<u>Test of Reading Comprehension</u> <u>and Language Achievement (English)</u>			
1. Reading Comprehension Part (RC)	1771	3.4	2.7
2. First Language Achievement Part (L1)	1771	3.0	2.8
3. Second Language Achievement Part (L2)	1771	6.5	3.2
4. Total Test (RCLA)	1771	12.9	7.2
<u>Writing Test (W)</u>	793	5.6	1.3
<u>Test of Arithmetic and Basic Algebra (A.A.)</u>	1627	16.7	9.4
<u>School Marks</u>			
1. English-Advanced Level Courses (E-A)	1317	65.6	11.7
2. English-General Level Courses (E-G)	779	63.2	10.8
3. Mathematics-Advanced Level Course (M-A)	943	68.0	14.3
4. Mathematics-General Level Course (M-G)	623	65.2	13.6

^aThis was the number of students drawn in the original sample. The difference between this number and the number of students taking the Test of Reading Comprehension and Language Achievement or the Test of Arithmetic and Basic Algebra is accounted for by unusable test records and absenteeism.

TABLE F.7

Correlation Coefficients and Sample Sizes for Correlation Coefficients
Among Achievement Tests and School Marks for SSGD Students

<u>Variable^a</u>	<u>RC</u>	<u>L1</u>	<u>L2</u>	<u>RCLA</u>	<u>W</u>	<u>A.A.</u>	<u>E-A</u>	<u>E-G</u>	<u>M-A</u>	<u>M-G</u>
RC		1771	1771	1771	790	1592	992	519	710	449
L1	55		1771	1771	790	1592	992	519	710	449
L2	51	51		1771	790	1592	992	519	710	449
RCLA	81	82	84		790	1592	992	519	710	449
W	51	49	50	61		723	453	270	328	216
A.A.	47	45	40	53	40		907	479	684	420
E-A	34	38	35	43	41	33		71	767	239
E-G	22	21	31	33	36	20	27		104	317
M-A	26	27	20	30	21	56	54	47		x ^b
M-G	09	13	16	17	25	39	43	37	x	

Note: Decimals have been omitted from the correlation coefficients, which appear below the diagonal. The size of the sample for each coefficient appears above the diagonal.

^aFor key to variable labels, see Table F.6

^bCells marked with an "X" involved fewer than 30 students; correlations are not reported for these combinations of variables.

APPENDIX G

TABLES OF STATISTICS REFERRED TO IN THE
SUBSECTION ON MARKING STANDARDS (CHAPTER 3, 2.2)
OF THE SSGD/SSHGD SURVEY
OF FRANCOPHONE STUDENTS

TABLE G. 1

Summary Statistics From the Multiple-Regression Analyses of Marks in Français Courses

Intercorrelations

<u>Variable^a</u>	<u>S.D.</u>	<u>B</u>	<u>S.E. of B</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
SSHGD Analysis							
1. School Marks	9.9			-			
2. Reading Comprehension Part	2.5	0.85	0.30	0.41	-		
3. First Language Achievement Part	2.1	0.32	0.35	0.29	0.34	-	
4. Second Language Achievement Part	2.3	0.35	0.32	0.32	0.37	0.26	-
5. Test de composition écrite	1.3	0.33	0.06	0.54	0.36	0.31	0.35

Multiple Correlation Coefficient = 0.60; N=151; n=12

SSGD Analysis - Advanced Level Course

1. School Marks	10.0			-			
2. Reading Compréhension Part	2.7	-0.28	0.32	0.28	-		
3. First Language Achievement Part	2.4	1.20	0.37	0.53	0.30	-	
4. Second Language Achievement Part	2.4	0.31	0.37	0.39	0.36	0.52	-
5. Test de composition écrite	1.2	0.38	0.07	0.60	0.53	0.48	0.41

Multiple Correlation Coefficient = 0.66; N=118; n=14

TABLE G.1 (continued)

Variable ^a	S.D.	B	S.E. of B	Intercorrelations			
				1	2	3	4
SSHGD Analysis							
SSGD Analysis - General Level Course							
1. School Marks	8.4			-			
2. Reading Comprehension Part	2.3	0.92	0.34	0.32	-		
3. First Language Achievement Part	1.7	1.17	0.47	0.31	0.11	-	
4. Second Language Achievement Part	1.7	0.68	0.47	0.21	0.06	0.17	-
5. Test de composition écrite	1.2	0.09	0.06	0.26	0.24	0.18	0.18

Multiple Correlation Coefficient = 0.56; N=102; n=13

Note: Statistics S.D., B--the raw "partial" score regression coefficient, S.D. of B, and the intercorrelations among the variables were derived from pooled within-school variation and covariation. N is the number of students; n the number of schools.

^aVariables 2,3 and 4 were scores on the three constituent parts of the Test de compréhension en lecture et de connaissance de la langue (français). Students with marks in more than one français course were assigned an average mark.

TABLE G.2

Summary Statistics From the Multiple-Regression Analyses of Marks in English/Anglais Courses

<u>Variable^a</u>	<u>S.D.</u>	<u>B</u>	<u>S.E. of B</u>	Intercorrelations		
				<u>1</u>	<u>2</u>	<u>3</u>
SSHGD Analysis						
1. School Marks	9.3			-		
2. Reading Comprehension Part	3.2	0.81	0.17	0.36	-	
3. Writing Exercise - Summary	1.6	0.08	0.03	0.21	0.13	-
4. Writing Exercise - Commentary	2.0	0.14	0.03	0.39	0.22	0.13
Multiple Correlation Coefficient = 0.50; N=241; n=12						
SSGD Analysis - Advanced Level Course						
1. School Marks	7.9			-		
2. Reading Comprehension Part	3.6	0.92	0.21	0.53	-	
3. Writing Exercise - Summary	1.7	0.01	0.05	0.28	0.41	-
4. Writing Exercise - Commentary	1.7	0.14	0.04	0.44	0.29	0.25
Multiple Correlation Coefficient = 0.61; N=81; n=13						

TABLE G.2 (continued)

Variable ^a	S.D.	B	S.E. of B	Intercorrelations		
				<u>1</u>	<u>2</u>	<u>3</u>
SSSGD Analysis - General Level Course						
1. School Marks	10.0			-		
2. Reading Comprehension Part	3.9	0.55	0.29	0.35	-	
3. Writing Exercise - Summary	1.7	0.09	0.07	0.18	0.15	-
4. Writing Exercise - Commentary	1.2	0.35	0.09	0.49	0.27	-0.01
Multiple Correlation Coefficient = 0.56; N=61; n=13						

Note: Statistics S.D., B - the raw score regression coefficient, S.E. of B, and the intercorrelations among the variables were derived from pooled-within school variation and covariation. N is the number of students; n the number of schools.

^aVariables 2, 3 and 4 were the three scores assigned student performance on the Test de connaissance de la langue (anglais). Students with marks in more than one English/anglais course were assigned an average mark.

TABLE G.3
Summary Statistics From the Regression Analyses of Marks in
Mathematics and Physics Courses

<u>Variable</u>	<u>S.D.</u>	<u>B</u>	<u>S.E. of B</u>	<u>r</u>	<u>N</u>	<u>n</u>
SSGD Analysis						
1. School Marks in Mathematics ^a	12.4					
2. <u>Test de rendement en mathématiques</u>	5.2	1.48	0.18	0.62	109	12
1. School Marks in Physics	9.8					
2. <u>Test de rendement en physique</u>	5.1	1.02	0.21	0.52	63	12
SSGD Analysis						
1. School Marks in Advanced Mathematics ^b	13.4					
2. <u>Test d'arithmétique et d'algèbre de base</u>	6.8	1.39	0.14	0.70	107	13
1. School Marks in General Mathematics ^c	10.6					
2. <u>Test d'arithmétique et d'algèbre de base</u>	3.8	0.09	0.39	0.03	51	10

TABLE G.3 (continued)

Note: Statistics S.D., B--the raw score regression coefficient, S.E. of B, and r--the correlation coefficient were derived from pooled within-school variation and covariation. N is the number of students, n the number of schools.

^aAverage of marks in the calculus course, and the functions and relations course.

^bMark in the Grade 12 "foundations" mathematics course.

^cMark in the Grade 12 "applications" mathematics course.

^dAverage of marks in any other Grade 12 mathematics courses.

TABLE G.4

Means, Standard Deviations and Sample Sizes for Distributions
of Achievement Test Scores and School Marks-SSGD Students

<u>Variable</u>	<u>N</u> <u>(656)^a</u>	<u>Mean</u>	<u>S.D.</u>
<u>Test de compréhension en lecture et</u> <u>de connaissance de la langue (français)</u>			
1. Reading Comprehension Part (RC)	460	6.8	2.8
2. First Language Achievement Part (L1)	460	2.9	2.5
3. Second Language Achievement Part (L2)	460	3.9	2.7
4. Total Test (TCLCL)	460	13.6	6.1
<u>Test de composition écrite (CE)</u>	183	6.5	1.4
<u>Test de connaissance de la langue (anglais)</u>			
1. Reading Comprehension Part (A-RC)	263	12.8	3.5
2. Writing Exercise - Summary Score (A-S)	261	3.7	1.8
3. Writing Exercise - Commentary Score (A-C)	261	4.9	2.1
<u>Test de rendement en mathématiques (MAT)</u>	114	14.6	5.8
<u>Test de rendement en physique (PAT)</u>	69	10.7	7.6
<u>School Marks</u>			
1. Français (F)	470	68.0	10.7
2. English/Anglais (E)	460	68.8	10.0
3. Mathematics (M)	370	67.7	13.9
4. Physics (P)	191	69.0	12.9

TABLE G.4 (continued)

a This was the number of students drawn in the original sample. The difference between this number and the number of students who took the Test de compréhension en lecture et de connaissance de la langue (français) is due primarily to absenteeism on the day the tests were administered.

TABLE G.5

Correlation Coefficients and Sample Sizes for Correlation
Coefficients Among Test Scores and School
Marks Obtained for SSGD Students

<u>Variables^a</u>	<u>RC</u>	<u>L1</u>	<u>L2</u>	<u>TCCL</u>	<u>CE</u>	<u>A-RC</u>	<u>A-S</u>	<u>A-C</u>	<u>MAT</u>	<u>PAT</u>	<u>F</u>	<u>E</u>	<u>M</u>	<u>P</u>
RC		460	460	460	183	256	253	253	112	68	347	346	280	140
L1	31		460	460	183	256	253	253	112	68	347	346	280	140
L2	39	43		460	183	256	253	253	112	68	347	346	280	140
TCCL	76	74	78		183	256	253	253	112	68	347	346	280	140
CE	36	41	39	52		115	112	112	32	^b 151	143	97		38
A-RC	39	31	38	46	08		259	259	-	X	210	245	133	55
A-S	5	7	17	12	11	16		261	-	X	209	243	131	55
A-C	24	22	23	30	33	29	-		-	X	209	243	131	55
MAT	40	20	38	40	13	-	-	-		36	70	68	109	69
PAT	44	16	34	41	-	X	X	X	88		34	45	64	63
F	36	37	36	47	51	31	13	19	26	33		355	252	103

TABLE G.5 (continued)

<u>Variables^a</u>	<u>RC</u>	<u>L1</u>	<u>L2</u>	<u>TCLCL</u>	<u>CE</u>	<u>A-RC</u>	<u>A-S</u>	<u>A-C</u>	<u>MAT</u>	<u>PAT</u>	<u>F</u>	<u>E</u>	<u>M</u>	<u>P</u>
E	36	31	30	42	39	36	14	39	47	46	62		258	124
M	31	22	28	36	24	25	15	13	56	52	48	46		179
P	41	8	22	32	30	22	17	27	60	54	56	53	72	

Note: Correlation coefficients appear below the diagonal, the size of the sample for each coefficient appears above the diagonal. Decimal points have been omitted from the correlation coefficients.

^aFor a key to the variables, see Table G.4

^bCells marked with a dash involve combinations of tests taken by no students.

^cCells marked with an X involved fewer than 30 students, hence results are not reported for these combinations of variables.

TABLE G.6

Means, Standard Deviations and Sample Sizes for Distribution
of Achievement Test Scores and School Marks Obtained on
SSGD Students

<u>Variable</u>	<u>N</u> <u>(705)^a</u>	<u>Mean</u>	<u>S.D.</u>
<u>Test de compréhension en lecture et de</u> <u>connaissance de la langue (français)</u>			
1. Reading Comprehension Part (RC)	514	5.3	3.0
2. Language Achievement Part (L1)	514	2.2	2.5
3. Second Language Achievement Part (L2)	514	2.4	2.6
4. Total Test (TCLCL)	514	9.8	6.5
<u>Test de composition écrite (CE)</u>			
	245	5.6	1.6
<u>Test de connaissance de la langue (anglais)</u>			
1. Reading Comprehension Part (A-RC)	175	10.5	4.8
2. Writing Exercise-Summary Score (A-S)	170	3.3	1.9
3. Writing Exercise-Commentary Score (A-C)	170	4.0	1.9
<u>Test d'arithmétique et d'algèbre de base (AA)</u>	291	10.7	9.7

TABLE G.6 (continued)

<u>Variable</u>	<u>N</u> <u>(705)^a</u>	<u>Mean</u>	<u>S.D.</u>
School Marks			
1. Français-Advanced Level Courses (F-A)	337	67.5	10.8
2. Français-General Level Courses (F-G)	268	64.5	10.7
3. English/Anglais-Advanced Level Courses (E-A)	314	68.4	10.0
4. English/Anglais-General Level Course (E-G)	265	66.0	10.5
5. Mathematics-Advanced Level Course (M-A)	246	69.1	14.5
6. Mathematics-General Level Course (M-G)	124	64.3	13.0

^aThis was the number of students drawn in the original sample. The difference between this number and the number of students who took the Test de compréhension en lecture et de connaissance de la langue (français) is due in large measure to absenteeism on the day the tests were administered.

TABLE G.7

Correlation Coefficients and Sample Sizes for Correlation Coefficients Among Achievement Tests and School Marks for SSGD Students

Variable ^a	<u>RC</u>	<u>L1</u>	<u>L2</u>	<u>ICLCL</u>	<u>CE</u>	<u>A-RC</u>	<u>A-S</u>	<u>A-C</u>	<u>AA</u>	<u>F-A</u>	<u>F-G</u>	<u>E-A</u>	<u>E-G</u>	<u>M-A</u>	<u>M-G</u>
RC		514	514	514	245	175	170	170	287	264	198	242	194	196	87
L1	42		514	514	245	175	170	170	287	264	198	242	194	196	87
L2	47	52		514	245	175	170	170	287	264	198	242	194	196	87
ICLCL	82	79	80		245	175	170	170	287	264	198	242	194	196	87
CE	53	54	49	65		77	75	75	141	118	102	111	93	90	43
A-RC	55	34	51	57	56		170	170	-	90	67	81	66	69	30
A-S	32	40	42	46	41	39		170	-	88	64	81	61	68	30
A-C	41	28	31	41	53	41	19		-	88	64	81	61	68	30
AA	51	38	42	55	42	_b	-	-		152	113	141	108	107	51
F-A	38	45	38	51	61	52	52	40	43		X	227	67	198	41
F-G	32	37	18	42	28	15	14	20	9	X ^c		55	173	34	67
E-A	30	38	30	41	53	43	31	30	29	70	70		X	166	42

TABLE G.7 (continued)

Variable ^a	<u>RC</u>	<u>L1</u>	<u>L2</u>	<u>TCLCL</u>	<u>CE</u>	<u>A-RC</u>	<u>A-S</u>	<u>A-C</u>	<u>AA</u>	<u>F-A</u>	<u>F-G</u>	<u>E-A</u>	<u>E-G</u>	<u>M-A</u>	<u>M-G</u>
E-G	21	17	10	23	22	40	21	27	29	51	49	X		48	67
M-A	31	44	38	46	48	38	36	41	59	57	51	60	34		X
M-G	19	02	7	14	21	38	26	16	12	46	41	48	12	X	

Note: Correlation coefficients appear below the diagonal; the size of the sample for each coefficient appears above the diagonal. Decimal points have been omitted from the correlation coefficients.

^aFor a key to the variable labels, see Table G.6.

^bCells marked with a dash involve combinations of tests taken by no students.

^cCells marked with an X involved fewer than 30 students, hence results are not reported for these combinations of variables.

APPENDIX H

TABLES OF STATISTICS REFERRED TO IN THE

REPORT OF RESULTS FROM THE

UNIVERSITY SURVEY

(CHAPTER 3, 3.1)

Number
Codes
for Universities

1. Brock
2. Carleton
3. Guelph
4. Lakehead
5. Laurentian
6. McMaster
7. Ottawa
8. Toronto
9. Trent
10. Waterloo
11. Windsor

TABLE H.1
Frequency Distribution of the Number of
Humanities/Arts Courses Completed by Students
From Each University

Number of Courses	University										
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>
13											
12							1				
11											
10								2			
9											
8								2			4
7						1	1	4		1	4
6			6				5	5		1	9
5		1	3		3	2	5	17		3	4
4		2	13	1	1	7	13	23		5	
3	4	8	10		1	6	7	34		14	4
2	4	15	13	1	4	19	12	79	3	23	28
1	12	21	40	15	16	23	19	98	9	42	39
0 ^a	4	1	2	2	1	5	1	7		3	7

TABLE H.1 (continued)

Number of Courses	University										
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>
Mean	1.3	1.8	2.3	1.1	1.8	2.0	3.0	2.3	1.3	1.9	2.5
S.D.	1.0	1.0	1.6	0.8	1.4	1.4	2.1	1.7	0.5	1.3	2.2
N ^b	24	48	87	19	26	63	64	271	12	92	46
N ^c	6	39	41	8	10	84	8	220	3	102	33

Note: No distinction was made in this tabulation between courses in terms of duration (one term, two terms, etc.), credit value, or hours of instructional time.

^aThis number was assigned to students who were shown to be enrolled in at least one Humanities/Arts course but who failed to receive a mark in the course(s).

^bNumber of students enrolled in at least one Humanities/Arts Course.

^cNumber of students enrolled in other courses but not enrolled in even one Humanities/Arts course.

TABLE H.2
Frequency Distribution of the Number of
Social Science Courses Completed by
Students From Each University

Number of Courses	University										
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>
9										3	1
8			1							4	
7			2							6	1
6			12			2		1		9	3
5		1	17			5	1	9	2	14	2
4		1	29	1	3	13	5	32	3	33	4
3	4	13	29	5	11	32	8	56	6	16	18
2	5	31	25	5	10	26	20	90	1	32	19
1	19	27	8	3	6	52	13	154	3	15	10
0 ^a	2	3	2			5	2	17		3	2

TABLE H.2 (continued)

Number of Courses	University										
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>
Mean	1.4	1.8	3.5	2.3	2.4	2.1	2.1	1.8	3.0	3.6	2.7
S.D.	0.8	0.9	1.6	0.9	0.9	1.3	1.1	1.1	1.3	2.0	1.7
N ^b	30	76	125	14	30	135	49	349	15	135	60
N ^c		11	3	13	6	12	23	142		59	19

Note: No distinction was made in this tabulation between courses in terms of duration (one term, two terms, etc.), credit value, or hours of instructional time.

^aThis number was assigned to students who were shown to be enrolled in at least one Social Science course, but who failed to receive a mark in the course(s).

^bNumber of students enrolled in at least one Social Science course.

^cNumber of students enrolled in other courses but not taking even one Social Science course.

TABLE H.3
Frequency Distribution of the Number of
Science Courses Completed by Students
From Each University

Number of Courses	University										
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>
14										1	
13										2	
12										7	
11											
10										8	
9			1							6	
8										5	
7								5		12	1
6			16					21		15	1
5	5		24	1	1	1		17		30	1
4	2	7	15	7		14	6	41	1	21	11
3	3	4	12	4	2	20	9	62		17	11
2	2	12	26	2	1	13	6	79	1	12	6
1	18	15	11	5	3	30	9	134	1	21	23
0 ^a		9	6	3	2	6	7	23	1	7	2

TABLE H.3 (Continued)

Number of Courses	University										
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>
Mean	2.1	1.7	3.5	2.5	1.8	2.1	1.9	2.3	1.8	4.8	2.3
S.D.	1.6	1.3	1.9	1.6	1.6	1.3	1.4	1.6	1.7	3.1	1.5
N ^b	30	47	111	22	27	84	37	382	4	164	56
N ^c		40	17	5	9	63	35	109	11	30	23

Note: No distinction was made in this tabulation between courses in terms of duration (one term, two terms, etc.), credit value, or hours of instructional time.

^aThis number was assigned to students who were shown to be enrolled in at least one Science course but who failed to receive a mark in the course(s).

^bNumber of students enrolled in at least one Science course.

^cNumber of students enrolled in other courses but not enrolled in even one Science course.

TABLE H.4
Frequency Distribution of the Number of
Professional Courses Completed by
Students From Each University

Number of Courses	University										
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>
10								6			
9								5		1	8
8		1									1
7										5	2
6			1		1					19	7
5		2	1	6	3		1	13		22	4
4		3	4		5	12	3	37		6	2
3		8	8	3	2	5	2	34		17	12
2		17	30	4	8	22	12	40		28	8
1	14	29	30	5	8	17	14	103	2	22	4
0 ^a		1	3	1		5	3	36		5	4

TABLE H.4 (continued)

Number of Courses	University										
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>
Mean	1.0	1.9	1.8	2.7	2.6	2.0	1.7	2.2	1.0	3.4	4.2
S.D.		1.3	1.1	1.8	1.5	1.2	1.2	2.1		2.1	2.8
N ^b	14	61	77	19	27	61	35	274	2	125	52
N ^c	16	26	51	8	9	86	37	217	13	69	27

Note: No distinction was made in this tabulation between courses in terms of duration (one term, two terms, etc.), credit value, or hours of instructional time.

^aThis number was assigned to students who were shown to be enrolled in at least one Professional course but who failed to receive a mark in the course(s).

^bNumber of students enrolled in at least one Professional course.

^cNumber of students enrolled in other courses but not enrolled in even one Professional course.

TABLE H.5

Frequency Distribution of the Number of
Humanities/Arts Courses Completed by
Students From Each Program Area

Number of Courses	Program Area			
	<u>Humanities/ Arts</u>	<u>Social Science</u>	<u>Science</u>	<u>Professional</u>
13	1			
12				
11				
10	2			
9	1			
8	4			
7	11			
6	25			
5	36	3		
4	55	10	2	2
3	65	19	5	3
2	29	97	31	23
1	2	115	138	46
0 ^a		11	11	5

TABLE H.5 (continued)

Number of Courses	Program Area			
	<u>Humanities/ Arts</u>	<u>Social Science</u>	<u>Science</u>	<u>Professional</u>
Mean	3.8	1.7	1.2	1.4
S.D.	1.8	0.9	0.6	0.8
N ^b	231	255	187	138
N ^c		132	284	79

Note: No distinction was made in this tabulation between courses in terms of duration (one term, two terms, etc.), credit value, or hours of instructional time.

^aThis number was assigned to students who were shown to be enrolled in at least one Humanities/Arts course but who failed to receive a mark in the course(s).

^bNumber of students enrolled in at least one Humanities/Arts course.

^cNumber of students enrolled in other courses but not enrolled in even one Humanities/Arts course.

TABLE H.6

Frequency Distribution of the Number of
Social Science Courses Completed by
Students From Each Program Area

Number of Courses	Program Area			
	<u>Humanities/ Arts</u>	<u>Social Science</u>	<u>Science</u>	<u>Professional</u>
9		4		
8		5		
7		9		
6		27		
5	1	44	4	2
4	7	69	32	6
3	20	127	33	18
2	65	77	73	49
1	75	24	139	72
0 ^a	9	1 ^b	20	6

TABLE H.6 (continued)

Number of Courses	Program Area			
	<u>Humanities/ Arts</u>	<u>Social Science</u>	<u>Science</u>	<u>Professional</u>
Mean	1.7	3.5	1.8	1.7
S.D.	0.9	1.6	1.2	1.0
N ^c	177	387	301	153
N ^d	54		170	64

Note: No distinction was made in this tabulation between courses in terms of duration (one term, two terms, etc.), credit value, or hours of instructional time.

^aThis number was assigned students who were enrolled in at least one Social Science course but who failed to receive a mark in the course(s).

^bStudents were classified into program areas on the basis of the courses in which they were enrolled, not on the basis of courses completed.

^cNumber of students enrolled in at least one Social Science course.

^dNumber of students enrolled in other courses but not enrolled in even one Social Science course.

TABLE H.7
Frequency Distribution of the Number of
Science Courses Completed by
Students in Each Program Area

Number of Courses	Program Area			
	<u>Humanities/ Arts</u>	<u>Social Science</u>	<u>Science</u>	<u>Professional</u>
14			1	
13			2	
12				
11			7	
10			8	
9			7	
8			5	
7			18	
6			51	2
5		1	59	20
4		5	97	23
3	1	10	106	27
2	12	46	66	36
1	61	120	28	61
0 ^a	14	21	16 ^b	15

TABLE H.7 (continued)

Number of Courses	Program Area			
	<u>Humanities/ Arts</u>	<u>Social Science</u>	<u>Science</u>	<u>Professional</u>
Mean	1.0	1.3	4.0	2.3
S.D.	0.6	0.9	2.3	1.5
N ^c	88	203	471	184
N ^d	143	184		33

Note: No distinction was made in this tabulation between courses in terms of duration (one term, two terms, etc.), credit value, or hours in instructional time.

^aThis number was assigned students who were shown to be enrolled in at least one Science course but who failed to receive a mark in the course(s).

^bStudents were classified into programs on the basis of the courses in which they were enrolled, not on the basis of courses completed.

^cNumber of students enrolled in at least one Science course.

^dNumber of students enrolled in other courses but not enrolled in even one Science course.

TABLE H.8
Frequency Distribution of the Number of
Professional Courses Completed by
Students in Each Program Area

Number of Courses	Program Area			
	<u>Humanities/ Arts</u>	<u>Social Science</u>	<u>Science</u>	<u>Professional</u>
10				6
9				14
8				2
7				7
6			3	25
5	1	4	11	36
4	5	14	27	26
3	4	18	31	38
2	11	34	81	43
1	22	83	130	13
0 ^a	4	19	28	7 ^b

TABLE H.8 (continued)

Number of Courses	Program Area			
	<u>Humanities/ Arts</u>	<u>Social Science</u>	<u>Science</u>	<u>Professional</u>
Mean	1.7	1.6	1.8	4.1
S.D.	1.2	1.2	1.3	2.4
N ^c	47	172	311	217
N ^d	184	215	160	

Note: No distinction was made in the tabulation between courses in terms of duration (one term, two terms, etc.), credit value, or hours of instructional time.

^aThis number was assigned students who were enrolled in at least one Professional course but who failed to receive a mark in the course(s).

^bStudents were classified into program areas on the basis of the courses in which they were enrolled, not on the basis of courses completed.

^cNumber of students enrolled in at least one Professional course.

^dNumber of students enrolled in other courses but not enrolled in even one Professional course.

APPENDIX I
TABLES OF STATISTICS FOR THE
STUDY OF FACTORS RELATED TO SCHOOL ACHIEVEMENT
(CHAPTER 3, 1.4)

TABLE I.1

Means and Standard Deviations of the Distributions of the Variables Included in the SSGD Study of Factors Associated with School Achievement				
<u>Variable</u>	<u>N</u>	<u>Mean</u>	<u>S.D.1</u>	<u>S.D.2</u>
Dependent Variables				
1. <u>Test of Reading Comprehension and and Language Achievement (English)</u>	1696	12.93	7.15	6.86
2. <u>Writing Test</u>	759	5.59	1.31	1.24
3. <u>Test of Arithmetic and Basic Algebra</u>	1528	16.75	9.43	8.99
4. <u>Future Plans^a</u>	1702	2.24	0.88	0.84
Independent Variables				
5. <u>Age (in years)</u>	1702	17.99	0.62	0.60
6. <u>Sex (1 = Female; 2 = Male)</u>	1702	1.48	0.50	0.49
7. <u>Language Spoken in the Home (2 = English; 1 = Other)</u>	1702	1.88	0.33	0.31
8. <u>Education of Father/Male Guardian</u>	1702	4.28	1.75	1.63
9. <u>Education of Mother/Female Guardian</u>	1702	4.44	1.64	1.56
10. <u>Occupation of Father/Male Guardian</u>	1702	3.89	1.46	1.35

TABLE I.1 (continued)

<u>Variable</u>	<u>N</u>	<u>Mean</u>	<u>S.D.2</u>	<u>S.D.2</u>
Independent Variables				
11. Occupation of Mother/Female Guardian (1 = Gainfully employed; 0 = Other)	1702	0.42	0.49	0.48
12. Total Credits	1702	28.78	2.17	2.07
13. Total Credits in English	1702	3.99	0.69	0.67
14. Total Credits in Mathematics	1702	3.69	0.99	0.94
15. Total Credits in Advanced Courses	1702	12.68	2.67	6.79
16. Total Credits in General Courses	1702	4.45	5.14	4.58

Note: Means and S.D.1 refer to the distribution of the variable over the total sample; S.D.2 is the standard deviation of the pooled within-school distribution for each variable.

^aFuture Plans was coded as follows: 3 = Continue in Secondary School; 2 = Go on to Postsecondary Study; 1 = Other Plans.

TABLE I.2

Coefficients of Correlation Among the Variables Included in the
SSGD Study of Factors Associated with School Achievement
(Coefficients as Based on the Pooled Within-school Variance and Covariance)

Variable ^a	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>	<u>12</u>	<u>13</u>	<u>14</u>	<u>15</u>	<u>16</u>
Dependent																
1	-	756	1525	1696								1696
2	57	-	690	759								759
3	51	38	-	1528								1528
4	35	31	49	-	1702							1702
Independent																
5	-30	-24	-27	-23	-											
6	-09	-21	08	03	12	-										
7	07	04	05	-03	-07	02	-									
8	21	16	20	17	-08	02	10	-								
9	17	08	14	15	-09	01	17	49	-							(1702) ^b
10	20	12	15	16	-13	-01	12	48	27	-						
11	-02	00	-01	02	-02	-01	06	00	12	-03	-					

TABLE I.2 (continued)

<u>Variable^a</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>	<u>12</u>	<u>13</u>	<u>14</u>	<u>15</u>	<u>16</u>
Dependent																
12	31	30	45	28	-20	-09	03	06	07	06	-04	-				
13	18	23	09	05	-13	-17	10	03	.04	04	02	23	-			
14	19	13	46	23	-18	-06	02	04	00	06	-02	43	12	-		
15	51	45	65	60	-30	-02	-02	25	19	23	01	44	17	33	-	
16	-38	-33	-53	-49	23	02	05	-21	-16	-20	-01	-24	05	-13	-86	-

Note: Coefficients of correlation are given below the diagonal, with decimal points omitted. The size of the sample on which each coefficient is based is given above the diagonal.

^aFor names of variables, see Table I.1.

^bAll coefficients for pairs of independent variables were based on a sample size of 1702.

TABLE I.3

Means and Standard Deviations of the Distributions of
 School Means on the Variables Included in the SSGD Study
 of Factors Related to School Achievement
 (Number of Schools = 51)

<u>Variable</u>	<u>Mean</u>	<u>S.D.</u>
Residuals (Observed Mean - Predicted Mean)		
1. <u>Test of Reading Comprehension and Language Achievement (English)</u>	0.00	1.49
2. <u>Writing Test</u>	0.00	0.38
3. <u>Test of Arithmetic and Basic Algebra</u>	0.00	2.36
4. Future Plans	0.00	0.21
Predicted Means		
5. <u>Test of Reading Comprehension and Language Achievement (English)</u>	12.84	1.08
6. <u>Writing Test</u>	5.57	0.15
7. <u>Test of Arithmetic and Basic Algebra</u>	16.78	1.72
8. Future Plans	2.25	0.11
Observed Means		
9. <u>Test of Reading Comprehension and Language Achievement (English)</u>	12.84	1.98
10. <u>Writing Test</u>	5.57	0.43
11. <u>Test of Arithmetic and Basic Algebra</u>	16.78	2.98
12. Future Plans	2.25	0.28
13. Age	18.00	0.15
14. Sex	1.49	0.11
15. Language Spoken in the Home	1.87	0.11

TABLE I.3 (continued)

<u>Variable</u>	<u>Mean</u>	<u>S.D.</u>
16. Education of Father/Male Guardian	4.27	0.64
17. Education of Mother/Female Guardian	4.40	0.54
18. Occupation of Father/Male Guardian	3.90	0.60
19. Occupation of Mother/Female Guardian	0.41	0.10
20. Total Credits	28.81	0.65
21. Total Credits in English	3.99	0.18
22. Total Credits in Mathematics	3.69	0.34
23. Total Credits in Advanced Subjects	12.64	3.51
24. Total Credits in General Subjects	4.41	2.41

TABLE I.4
Coefficients of Correlations Among School Means on Variables
Included in the SSGD Study of Factors Related to School Achievement
(Number of Schools = 51)

Variable	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>	<u>12</u>	<u>13</u>	<u>14</u>	<u>15</u>	<u>16</u>	<u>17</u>	<u>18</u>	<u>19</u>	<u>20</u>	<u>21</u>	<u>22</u>	<u>23</u>	<u>24</u>	
1	-																								
2	52	-																							
3	11	45	-																						
4	26	16	46	-																					
5	16	20	18	34	-																				
6	20	16	18	26	96	-																			
7	08	05	04	19	86	79	-																		
8	14	16	16	39	93	82	93	-																	
9	84	50	44	38	67	58	53	61	-																
10	53	94	16	23	51	49	32	42	68	-															
11	40	11	82	48	64	60	61	66	65	31	-														
12	25	19	42	93	64	53	52	70	54	35	63	-													
13 ^a	-16	-22	-12	-36	-79	-81	-63	-65	-56	-48	-46	-54	-												
14 ^a	-25	07	-18	-01	-40	-57	-29	-20	-40	-14	-31	-09	44	-											
15 ^a	02	09	01	-06	60	61	43	35	34	29	26	09	-62	-56	-										

TABLE I.4 (continued)

Variable	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>	<u>12</u>	<u>13</u>	<u>14</u>	<u>15</u>	<u>16</u>	<u>17</u>	<u>18</u>	<u>19</u>	<u>20</u>	<u>21</u>	<u>22</u>	<u>23</u>	<u>24</u>
16 ^a	26	34	38	49	71	60	48	73	59	51	58	67	-34	-09	17	-	-	-	-	-	-	-	-	-
17 ^a	14	21	21	32	82	72	57	75	55	43	49	55	-51	-32	51	81	-	-	-	-	-	-	-	-
18 ^a	13	22	30	49	73	55	56	78	50	39	56	69	-38	00	28	88	77	-	-	-	-	-	-	-
19 ^a	14	00	-06	12	24	20	22	30	24	07	08	21	-16	-07	20	16	34	30	-	-	-	-	-	-
20	-31	-36	-25	-09	-31	-31	-17	-21	-41	-43	-30	-15	24	27	-38	-35	-32	-28	03	-	-	-	-	-
21 ^a	-13	-05	-19	-24	23	36	09	04	03	08	-10	-17	-18	05	19	-09	06	-14	03	01	-	-	-	-
22 ^a	-03	-17	-14	-10	51	50	86	63	25	02	39	18	-35	-28	29	03	17	15	14	-01	05	-	-	-
23	36	14	17	46	50	46	47	55	55	28	40	58	-34	-14	07	45	44	43	14	-27	04	28	-	-
24	-40	-13	-28	-52	29	-25	-28	-37	-46	-20	-39	-55	13	05	11	-38	-23	-35	-10	21	10	-15	-82	-

Note: For names of variables, see Table I.3. Decimal points have been omitted from the correlation coefficients.

^a Variables used in the multiple regression equation to obtain predicted school means on the tests of achievement and on Future Plans.

TABLE I.5

Means and Standard Deviations of the Distributions of the Variables Included in the SSHGD Study of Factors Associated with School Achievement

<u>Variables</u>	<u>N</u>	<u>Mean</u>	<u>S.D.1</u>	<u>S.D.2</u>
Dependent Variables				
1. <u>Test of Reading Comprehension and Language Achievement (English)</u>	2085	17.67	6.91	6.64
2. <u>Writing Test</u>	535	6.44	1.25	1.15
3. <u>Mathematics Achievement Test</u>	788	17.12	5.92	5.52
4. <u>Future Plans (5 = Go on to Postsecondary Study; 4 = Other)</u>	2094	4.81	0.39	0.38
Independent Variables				
5. <u>Age (in years)</u>	2094	18.79	0.63	0.60
6. <u>Sex (1 = Female; 2 = Male)</u>	2094	1.50	0.50	0.49
7. <u>Language Spoken in the Home (2 = English; 1 = Other)</u>	2094	1.89	0.31	0.28
8. <u>Education of Father/Male Guardian</u>	2094	4.71	1.76	1.67
9. <u>Education of Mother/Female Guardian</u>	2094	4.72	1.59	1.52
10. <u>Occupation of Father/Male Guardian (Blisshen Scale)</u>	2094	4.24	1.55	1.46

TABLE I.5 (continued)

<u>Variables</u>	<u>N</u>	<u>Mean</u>	<u>S.D.1</u>	<u>S.D.2</u>
Independent Variables				
11. Occupation of Mother/Female Guardian (1 = Gainfully Employed; 0 = Other)	2094	0.40	0.49	0.48
12. Total Credits	2094	35.30	2.40	2.17
13. Credits in English	2094	5.09	0.97	0.91
14. Credits in Mathematics	2094	5.46	1.57	1.49
15. Number of SSHGD Courses in Languages	2094	1.46	0.95	0.90
16. Number of SSHGD Courses in History/ Geography/Social Science	2094	0.85	0.90	0.85
17. Number of SSHGD Courses in Mathematics/Science	2094	1.73	1.20	1.10
18. Number of Other SSHGD Courses	2094	1.72	1.08	0.89
19. Total Number of SSHGD Courses	2094	5.71	1.15	1.06

Note: The Mean and S.D.1 refer to the distribution of the variable over the total sample; S.D.2 is the standard deviation of the pooled within school distribution for each variable.

TABLE I.6

Coefficients of Correlation Among the Variables Included in the
SSHGD Study of Factors Associated with School Achievement
(Coefficients are Based on the Pooled Within-School Variances and Covariances)

<u>Variable^a</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>	<u>12</u>	<u>13</u>	<u>14</u>	<u>15</u>	<u>16</u>	<u>17</u>	<u>18</u>	<u>19</u>
Dependent																			
1	-	532	785	2085											2085
2	42	-	190	535											535
3	36	22	-	788											788
4	09	13	08	-	2094										2094
Independent																			
5	-24	-18	-18	-03	-														
6	-11	-18	12	-05	14	-													
7	13	05	03	-08	-10	01	-												
8	11	09	-01	08	-08	01	14	-											
9	14	16	08	05	-10	-04	20	51	-										
10	12	12	07	04	-09	02	14	50	33	-									
11	00	06	-04	01	-02	-05	00	-01	10	-04	-								
12	20	24	16	19	-05	-11	-04	03	05	02	-01	-							
13	12	19	-06	01	01	-23	03	01	03	02	05	17	-						

(2094)^b

TABLE I.6 (continued)

Variable ^a	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>	<u>12</u>	<u>13</u>	<u>14</u>	<u>15</u>	<u>16</u>	<u>17</u>	<u>18</u>	<u>19</u>
Independent																			
14	10	-01	36	14	-07	18	-06	-01	03	03	-03	35	-15	-					
15	25	32	-01	04	-11	-37	03	02	02	04	00	16	49	-31	-				
16	-10	00	-20	-07	08	-05	08	-01	00	-02	02	-09	16	-47	09	-			
17	02	-13	31	08	00	21	-08	01	01	-01	-04	14	-32	52	-44	-50	-		
18	01	01	05	15	-04	11	-04	01	-01	02	-02	26	-21	53	-32	-39	19	-	
19	16	16	21	19	-06	-04	-03	04	01	02	-05	48	03	34	20	03	42	45	-

Note: Coefficients of correlation appear below the diagonal, with decimal points omitted. The size of the sample on which each coefficient is based is given above the diagonal.

^aFor names of the variables, see Table I.5.

^bAll coefficients for pairs of independent variables were based on a sample size of 2094.

TABLE I.7

Means and Standard Deviations of the Distributions of
School Means on the Variables Included in the SSHGD Study
of Factors Related to School Achievement
(Number of Schools = 52)

<u>Variable</u>	<u>Mean</u>	<u>S.D.</u>
Residuals (Obtained Mean-Predicted Mean)		
1. <u>Test of Reading Comprehension and Language Achievement (English)</u>		
2. Future Plans	0.00	0.08
Predicted Means		
3. <u>Test of Reading Comprehension and Language Achievement (English)</u>	17.58	1.22
4. Future Plans	4.82	0.03
Observed Means		
5. <u>Test of Reading Comprehension and Language Achievement (English)</u>	17.58	2.05
6. Future Plans	4.82	0.08
7. <u>Writing Test</u> ^a	6.41	0.58
8. <u>Mathematics Achievement Test</u>	17.03	2.37
9. Age	18.80	0.20
10. Sex	1.52	0.14
11. Language Spoken in the Home	1.81	0.17
12. Education of Father/Male Guardian	4.66	0.59
13. Education of Mother/Female Guardian	4.69	0.53
14. Occupation of Father/Male Guardian	4.21	0.57
15. Occupation of Mother/Female Guardian	0.40	0.09

TABLE I.7 (continued)

<u>Variable</u>	<u>Mean</u>	<u>S.D.</u>
16. Total Credits	35.32	1.03
17. Total Credits in English	5.09	0.38
18. Total Credits in Mathematics	5.56	0.65
19. Number of SSHGD Courses in Languages	1.39	0.34
20. Number of SSHGD Courses in History/ Geography/Social Science	0.82	0.30
21. Number of SSHGD Courses in Mathematics/Science	1.77	0.54
22. Number of Other SSHGD Courses	1.76	0.72
23. Total Number of SSHGD Courses	5.73	0.45

^aThe number of schools for the Writing Test was 50.

TABLE I.8

Coefficients of Correlation Among School Means on Variables
Included in the SSHGD Study of Factors Related to School Achievement

Variable	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>	<u>12</u>	<u>13</u>	<u>14</u>	<u>15</u>	<u>16</u>	<u>17</u>	<u>18</u>	<u>19</u>	<u>20</u>	<u>21</u>	<u>22</u>	<u>23</u>
1	-																						
2	35	-																					
3	12	-23	-																				
4	-25	-04	10	-																			
5	81	12	68	-13	-																		
6	24	94	18	31	07	-																	
7	25	09	45	-29	46	-02	-																
8	21	09	23	20	29	16	17	-															
9 ^a	-27	15	-81	10	-68	18	-43	-22	-														
10 ^a	-17	08	-32	37	-31	20	-36	03	26	-													
11 ^a	21	-23	83	-37	65	-35	49	05	-64	-45	-												
12 ^a	27	-09	57	03	54	-07	32	23	-48	-12	53	-											
13 ^a	29	-18	73	-06	65	-19	37	18	-48	-32	77	80	-										
14 ^a	24	-15	59	07	53	-12	25	16	-44	-08	54	91	95	-									
15 ^a	-18	-03	-18	08	-24	00	00	-13	15	02	-22	-15	-18	-21	-								

TABLE I.8 (continued)

Variable	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>	<u>12</u>	<u>13</u>	<u>14</u>	<u>15</u>	<u>16</u>	<u>17</u>	<u>18</u>	<u>19</u>	<u>20</u>	<u>21</u>	<u>22</u>	<u>23</u>
16	-10	20	09	45	-02	34	-16	11	02	14	-12	-06	-06	00	04	-	-	-	-	-	-	-	-
17 ^a	-22	-02	62	12	21	02	37	13	-36	-13	36	10	21	12	07	24	-	-	-	-	-	-	-
18 ^a	-28	-14	20	88	-08	18	-29	12	04	45	-15	-13	-06	-03	-02	39	14	-	-	-	-	-	-
19	18	12	25	-29	28	02	29	19	-31	-10	27	38	28	31	21	-02	26	-37	-	-	-	-	-
20	05	-11	19	-39	15	-24	19	16	-23	-16	34	40	33	34	-04	-13	05	-43	44	-	-	-	-
21	-08	16	-16	48	-15	33	-29	22	19	51	-34	-07	-24	-08	06	39	-08	48	-05	-18	-	-	-
22	-04	-09	07	23	01	00	-08	-18	03	-19	-01	-19	-02	-12	-16	04	01	27	-57	-46	-45	-	-
23	02	07	25	47	16	23	-12	24	-13	-14	01	19	11	18	-05	44	16	44	10	07	33	30	-

Note: For names of variables, see Table I.7; decimal points have been omitted from the correlation coefficients.

^aVariable used in the multiple regression equation to obtain predicted school means on the Test of Reading Comprehension and Language Achievement (English) and on Future Plans.

